BULLETIN

OF THE

AMERICAN MATHEMATICAL SOCIETY

EDITED BY

EARLE R. HEDRICK

DAVID R. CURTISS WILLIAM R. LONGLEY

WITH THE ASSISTANCE OF

HAROLD T. DAVIS HARRY W. KUHN

THEOPHIL H. HILDEBRANDT DERRICK N. LEHMER

JOHN R. KLINE HENRY L. RIETZ

CAROLINE E. SEELY

VOLUME XXXVIII, NUMBER 4

APRIL, 1932

PUBLISHED BY THE SOCIETY

MENASHA, WIS., AND NEW YORK

1932

PUBLISHED MONTHLY

Whole No. 391 $9.00 a Year

Entered as second class matter April 24, 1926, at the post office at Menasha, Wis., under the act of August 24, 1912.

AMERICAN MATHEMATICAL SOCIETY

Mail Address: 501 West 116th Street, New York, N.Y., for the Secretary, Treasurer, Librarian, and Clerk.

Office: Room 413, Physics Building, Columbia University, New York City.

President: Dean L. P. Eisenhart, Princeton University, Princeton, N.J.

Secretary: Dean R. G. D. Richardson, Brown University, Providence, R.I.

Associate Secretaries: Professor Tomlinson Fort, Lehigh University, Bethlehem, Pa.; Professor M. H. Ingraham, University of Wisconsin, Madison, Wis.; Dean T. M. Putnam, University of California, Berkeley, Calif.

Treasurer: Professor G. W. Mullins, Barnard College in Columbia University, New York City.

Librarian: Professor R. C. Archibald, Brown University, Providence, R.I.

Clerk: Dr. Caroline E. Seely, 413 Physics Building, Columbia University, New York City.

Editor-in-Chief of the Bulletin: Professor E. R. Hedrick, University of California at Los Angeles, Los Angeles, Calif.

Managing Editor of the Transactions: Professor R. D. Carmichael, University of Illinois, Urbana, Ill.

Chairman, Colloquium Editorial Committee: Professor R. L. Moore, University of Texas, Austin, Tex.

Managing Editor of the American Journal of Mathematics: Professor Abraham Cohen, Johns Hopkins University, Baltimore, Md.

Representatives of the Society in the Division of Physical Sciences of the National Research Council: Professors James Pierpont, D. R. Curtiss, and E. R. Hedrick.

Representatives on the Council of the American Association for the Advancement of Science: Professors Louis Ingold and M. H. Ingraham.

Society Visiting Lecturer for 1931-32: Professor R. L. Moore, University of Texas, Austin, Tex.

Colloquium Speaker for 1932: Professor J. F. Ritt, Columbia University, New York City.

Chairman, Committee on Arrangements for the Semi-Centennial: Professor T. S. Fiske, Columbia University, New York City.

THE CALCULUS OF VARIATIONS AND THE QUANTUM THEORY*

BY G. A. BLISS

1. *Introduction.* Several years ago an insistent curiosity caused me to forsake temporarily fields of mathematical interest more legitimate for me in order to find out if possible the character of the new quantum theories in which my physicist neighbors and some of my mathematical friends seemed to find so much of scientific interest and excitement. I was entirely ignorant of the theory at that time and was greatly surprised to find that my own specialty, the calculus of variations, through the media of mechanical and optical conceptions, had played an important role in its development. The quantum theory was then in an exceedingly fluent stage, and though the unanswered mathematical questions which presented themselves were numerous and significant it seemed to me useless for mathematicians to spend large amounts of time in endeavoring to perfect mathematical details when some new paper might at short notice cause fundamental changes in the whole structure of the theory. Every one who has followed the development of the quantum theory would agree, I think, that this impression was justified. At the present time the situation seems somewhat different. I have heard assurances from various reliable sources that the leaders in the promotion of quantum mechanics, men of wide physical experience and of daring in mathematical exploration distinctly beyond that of the average pure mathematician who takes his rules so seriously, have reached something like agreement. These statements should perhaps not be accepted too readily, as they sound like others that have been made in the past and which have afterward needed modification. But to the seeker after knowledge, less experienced in this field like myself, it also seems that an equilibrium has approx-

* Retiring Address of the Vice-President of Section A of the American Association for the Advancement of Science, delivered at New Orleans, December 29, 1931, at a joint meeting of this Society, the American Physical Society, and Section A of the A.A.A.S.

imately been reached in the fundamentals of the quantum theory, and that pure mathematicians could therefore reasonably endeavor to contribute in a valuable way to the development and perfection of its mathematical structure.

In the following pages I do not intend to attempt to point out mathematical defects in the quantum theory, though it seems clear that such defects exist. They appear for the most part in the generalities and less often in the special problems which have so far been successfully studied. My endeavor here will rather be to present some of the mathematical aspects of the theory to those of my mathematical colleagues who have not yet delved in the field. They will agree with me, I think, that there is no other applied mathematical domain whose development has involved a wider variety of mathematical techniques.

This paper was written originally as a retiring address for a chairman of the mathematicians in Section A of the Association. I was somewhat perturbed when I found that it had been allotted to a joint meeting of mathematicians and physicists. To those of the latter class who are here present I can only express my sympathy. It will be impossible for me to enlighten them with regard to quantum mechanics, and my mathematics may seem to them too technical. My only comfort is that it may possibly not seem more so than that of their own colleagues who specialize in the quantum theory.

2. *Normal Congruences and Optical Fields.* One of the simplest problems of the calculus of variations is that of finding in the class of arcs

$$y = y(x), \quad z = z(x), \qquad (x_1 \leqq x \leqq x_2),$$

joining two fixed points in xyz-space, one which minimizes the length integral

$$I = \int_{x_1}^{x_2} (1 + y'^2 + z'^2)^{1/2} dx.$$

This is a relatively simple problem concerning whose solution we know much intuitively. The possible minimizing curves are straight lines

$$y = ax + b, \quad z = cx + d, \tag{1}$$

the totality of which depends upon four constants a, b, c, d. *A congruence of lines* is a two-parameter family selected out of this four-parameter family by setting a, b, c, d equal to functions of two other parameters u, v. Geometers have long known that the lines of such a congruence will not in general have a surface which they all cut at right angles. If, however, there is one surface cutting all the lines of the congruence orthogonally, then there is a whole family of such surfaces given by an equation of the form

$$S(x, y, z) = \text{constant}. \tag{2}$$

The congruence is in that case called a *normal congruence*. Such congruences of lines have many interesting properties. For example the segments of lines of the congruence intercepted by two of the surfaces (2) are all equal in length, and the congruence has so-called focal surfaces to which all of the lines of the congruence are tangent. But the property which interests us most here is that a normal congruence may be determined by its family (2) of orthogonal surfaces as well as by its lines. Not every family (2) has a congruence of straight lines cutting it at right angles. The characteristic property of those families of surfaces which do belong to a normal congruence is that for such families the function $S(x, y, z)$ can always be selected so that it satisfies the differential equation

$$S_x^2 + S_y^2 + S_z^2 = 1,$$

where S_x, S_y, S_z are the partial derivatives of S with respect to x, y, and z.

All this will seem rather familiar to my physicist hearers because the principal business of the theory of geometrical optics is with normal congruences. The straight lines are the light rays, the orthogonal surfaces are the wave fronts, and the focal surfaces are the caustic surfaces of optical theory. An interesting property of an optical field of rays for the physicist is of course that it retains its character after a reflection or a refraction. Sir William Hamilton devoted one of his early published papers* to the study of the properties of such optical fields. It was the

* *Theory of systems of rays*, Transactions of the Royal Irish Academy, vol. 15 (1824), pp. 69–173; vol. 16, Part I, pp. 3–61, Part II, pp. 93–125; vol. 17, pp. 1–144.

more than 300-page effort of a youthful writer, and probably unnecessarily long, but it contains the source of the ideas which later influenced profoundly the development of the calculus of vaaritions and the theory of analytical mechanics.

3. *Generalizations in the Calculus of Variations.* It has been found that properties analogous to those of normal congruences are possessed by the minimizing curves of far more general problems of the calculus of variations than that of determining the shortest distance. When we try to minimize an integral

$$I = \int_{x_1}^{x_2} f(x, y_1, y_2, y_1', y_2')dx$$

in the class of arcs

$$y_1 = y_1(x), \quad y_2 = y_2(x), \qquad (x_1 \leqq x \leqq x_2)$$

joining two fixed points in xy_1y_2-space we find that the only possible minimizing curves are the solutions of the differential equations

$$\frac{df_{y_k'}}{dx} = f_{y_k}, \qquad (k = 1, 2). \tag{3}$$

I use the notations x, y_1, y_2 instead of x, y, z because then all of our equations can be readily interpreted for the $(r+1)$-dimensional case, instead of the 3-dimensional case, by simply changing the range $k=1, 2$ to $k=1, \cdots, r$. Meanwhile, for the space of points (x, y_1, y_2) we can use the language of ordinary 3-dimensional geometry. The differential equations (3) are of the second order and their solutions, which are called *extremals* in the calculus of variations, form a family of curves whose equations, like those of the straight lines (1), contain four arbitrary constants.

It has been found, following Hamilton and probably earlier writers also, that the equations (3) take a very pleasing form when new variables x, y_k, z_k are introduced in place of the original variables x, y_k, y_k' by means of the equations

$$z_k = f_{y_k'}(x, y, y'), \qquad (k = 1, 2).$$

Let us define a Hamiltonian function H by means of the equation

$$H(x, y, z) = y_k' f_{y_k'} - f,$$

it being understood here as elsewhere in this paper that a term containing the same subscript k twice stands for the sum of the terms of that form for the various special values in the range of k, as in tensor analysis. It can be proved that the equations of the extremals now take the form

$$\frac{dy_k}{dx} = H_{z_k}, \quad \frac{dz_k}{dx} = -H_{y_k}, \qquad (k = 1, 2). \tag{4}$$

The variables x, y_k, z_k are called *canonical variables*, and the equations (4) are the *canonical equations* of the extremals. Equations (4) may also be written in the form

$$\frac{dy_k}{dx} = (y_k, H), \quad \frac{dz_k}{dx} = (z_k, H), \qquad (k = 1, 2), \tag{5}$$

where the brackets on the right are so-called *Poisson brackets* defined for two arbitrary functions f and g of x, y_k, z_k by the equation

$$(f, g) = \frac{\partial f}{\partial y_k}\frac{\partial g}{\partial z_k} - \frac{\partial f}{\partial z_k}\frac{\partial g}{\partial y_k}. \tag{6}$$

For the fields of geometrical optics the equation which is satisfied when the direction $dx:dy:dz$ in xyz-space is at right angles with a direction $1:y':z'$ is the well known orthogonality condition

$$dx + y'\,dy + z'\,dz = 0.$$

In the calculus of variations it is found that this equation must be replaced by the equation

$$(f - y_k' f_{y_k'})dx + f_{y_k'}dy_k = 0.$$

A *direction* $dx:dy_1:dy_2$ *is transversal* to the direction $1:y_1':y_2'$ at the point (x, y) if it satisfies this equation; and a *surface is transversal* to $1:y_1':y_2'$ at (x, y) if all the directions $dx:dy_1:dy_2$ tangent to the surface at (x, y) satisfy the equation.

We are now in possession of data sufficient to justify a generalization of the notion of a normal congruence of straight lines. Instead of a two-parameter family of straight lines we consider a two-parameter family of the extremal curves which satisfy the differential equations (3) or (4). Such a congruence

of curves is called a Mayer family, after the name of its discoverer, if there is a single surface which cuts every member of the family transversally. It is provable, then, that there is a single infinity of such surfaces, the so-called transversal surfaces of the Mayer family, defined by an equation of the form

$$S(x, y_1, y_2) = \text{constant},$$

and analogous to (2). Not every such set of surfaces is the set of transversal surfaces of a Mayer family. In order that this property may be possessed by the surfaces it is necessary and sufficient that they should be definable by an equation $S(x, y) =$ constant for which S satisfies the partial differential equation

$$S_x + H(x, y, S_y) = 0.$$

This is the famous Hamilton-Jacobi partial differential equation of the calculus of variations. It has many interesting relationships with the extremals besides the one which has just been described.

We need one further notion from the calculus of variations in order to complete the list of those upon which the development of the quantum theory has depended. The variables x, y_k, z_k introduced above are not the only ones for which the differential equations of the extremals have the canonical form. Transformations

$$(7) \quad x = x(X, Y, Z), \quad y_k = y_k(X, Y, Z), \quad z_k = z_k(X, Y, Z),$$

which take x, y_k, z_k into new variables X, Y_k, Z_k and which transform the canonical differential equations into others of the same form are called *canonical transformations*. Such transformations can be constructed in many different ways but the most important ones for the quantum theory are those which replace (x, y_k, z_k) by new variables $(X, Y_k, Z_k) = (x, w_k, J_k)$ by means of equations of the form

$$(8) \qquad x = x, \quad z_k = S_{y_k}(x, y, J), \quad w_k = S_{J_k}(x, y, J),$$

where $S(x, y, J)$ is a solution of the differential equation

$$S_x + H(x, y, S_y) = E(J),$$

involving the variables J_i as parameters. The last set of equations (8) must be solved for the y_k and substituted in the second

in order to secure the transformation in the form (7). The important property of the transformation is that in terms of the variables x, w_k, J_k the differential equations of the extremals take the very simple form

$$\frac{dw_k}{dx} = E_{J_k}, \quad \frac{dJ_k}{dx} = 0,$$

and the equations of the extremals are

$$w_k = E_{J_k}(J)t + \alpha_k, \quad J_k = \beta_k, \tag{9}$$

where the α_k's and β_k's are constants.

The results from the calculus of variations mentioned in the preceding paragraphs have been described for the 3-dimensional space of points (x, y_1, y_2). But the equations which have been written are valid for the $(r+1)$-dimensional space of points $(x, y_1, \cdots, y_r)$ if we change the range of the indices k from 1, 2 to 1, $\cdots$, r. The number of parameters in the equations of the family of extremals is of course $2r$ instead of 4, and the Mayer families for $(r+1)$-dimensional space contain r parameters instead of only 2.

4. *Applications of the Calculus of Variations.* The calculus of variations appears in the quantum theory through its applications in mechanics. The equations of motion of a particle of mass m moving in a field of force with the potential function $V(x, y, z)$ are the equations

$$\frac{dmx'}{dt} = -V_x, \quad \frac{dmy'}{dt} = -V_y, \quad \frac{dmz'}{dt} = -V_z, \tag{10}$$

where the primes indicate derivatives with respect to t. According to the well known principle of Hamilton they are the differential equations of the extremals of the integral

$$I = \int_{t_1}^{t_2} (T - V)dt, \tag{11}$$

where $T = m(x'^2 + y'^2 + z'^2)/2$ is the kinetic energy of the moving particle. Thus all of the results of the calculus of variations described in the preceding section, with suitable changes in notation, find their applications in mechanics. It is not neces-

sary to review them in detail for this special case. It should be mentioned, however, that a similar formulation of Hamilton's principle holds for n particles instead of one, and for any auxiliary variables q_i in terms of which the coordinates of the particles may be expressed. The variables t, q_k, q_k' replace the x, y_k, y_k' of the preceding section. The kinetic energy T is a homogeneous quadratic form in the derivatives q_k' with coefficients functions of the q_k, and the canonical variables are t, q_k, p_k, where the momenta p_k are defined by the equations $p_k = T_{q_k'}$. The Hamiltonian function $H(t, p, q)$ turns out to be the total energy $H = T + V$ of the system and is equal to a constant E for the motion of every system.

This is a familiar basis for the theory of Newtonian analytical mechanics, but the modification of equations (10) which arises in the theory of relativity should also be mentioned. In the non-relativistic case the mass m is a constant, but in the relativity theory it varies in accordance with the law $m = m_0/(1 - v^2/c^2)^{1/2}$, where m_0 is the so-called rest-mass, v is the velocity of the particle, and c the velocity of light. The equations (10) are no longer the differential equations of the extremals of the integral (11) when T has the value described above. They will retain this property, however, if we replace T by the function

$$T_1 = m_0c^2[1 - (1 - v^2/c^2)^{1/2}],$$

as can readily be verified. The Hamiltonian function for this relativity case turns out to have the value $H = T_2 + V$, where the kinetic energy T_2 is now given by the formula

$$T_2 = m_0c^2[(1 - v^2/c^2)^{-1/2} - 1].$$

There are two further principles, important for the purposes of this paper, which are phrased in terms of concepts of the calculus of variations. The principle of least action in mechanics says that every trajectory of a particle of mass m, moving with total energy E in a field of force with the potential function V, is an extremal of the integral

$$\begin{aligned} I &= \int_{\tau_1}^{\tau_2} [2m(E - V)(x_\tau^2 + y_\tau^2 + z_\tau^2)]^{1/2} d\tau \\ &= \int_{s_1}^{s_2} [2m(E - V)]^{1/2} ds, \end{aligned} \tag{12}$$

where τ is a parameter in terms of which the coordinates x, y, z of the trajectory are to be expressed and s is the length of arc. The time t at which the particle arrives at the place defined on the trajectory by a value of τ is expressible by the equation

$$t = t_1 + \int_{\tau_1}^{\tau} [m(x_\tau^2 + y_\tau^2 + z_\tau^2)/(2(E - V))]^{1/2} d\tau.$$

Fermat's principle in geometrical optics has a form much like that of this mechanical principle of least action. It says that rays of light in a medium in which the velocity of light at a point (x, y, z) is $u(x, y, z)$ are extremals of the integral

$$(13)\quad I = C \int_{\tau_1}^{\tau_2} (1/u)(x_\tau^2 + y_\tau^2 + z_\tau^2)^{1/2} dz = C \int_{s_1}^{s_2} (1/u) ds,$$

where C is an arbitrary constant whose value does not affect the determination of the extremals in any way. The two integrals (12) and (13) are identical if

$$(14)\qquad u = C/[2m(E - V)]^{1/2}.$$

Hence the dynamical trajectories of a particle moving with total energy E are the same as the paths of light in a medium where the velocity of light at each point is given by the equation (14).

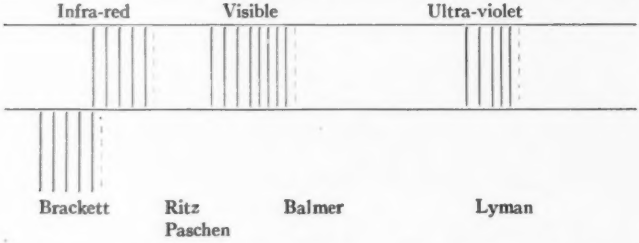
John Bernoulli used an analogy similar to this in 1697 when he determined the curve down which a particle would fall under the action of gravity from one fixed point to another in the shortest time. He found that the time of descent was given, except for a constant factor, by the integral

$$\int_{s_1}^{s_2} z^{-1/2} ds.$$

Hence the curve of quickest descent must be a light-ray in a medium where $u = Cz^{1/2}$. By dividing the medium into horizontal layers and applying the law of refraction he was able to find the differential equation of the minimizing curve and to show that the curve of quickest descent is an inverted cycloid. The paper in which he found this result is often designated as the origin of the calculus of variations.

5. *The Quantum Theory of Bohr.* In my effort to give my fellow mathematicians some idea of the role which the calculus of variations has played in the development of the quantum theory I do not hesitate to begin with the early theory of Bohr, even though it has been superseded in recent years by quite different mathematical mechanisms. It still affords a convenient picture by means of which the physicist correlates a considerable part of the data in which the quantum theory had its source, and with the theory of Schrödinger it forms the chapters of the development of the quantum theory in which the calculus of variations had its greatest influence. Furthermore I believe that a knowledge of these earlier stages is essential to an understanding of quantum mechanics in its more recent forms.

To the mathematicians I may explain that one of the first problems of the quantum theory, and one of the least complicated illustrations of it, is the theory of the spectrum of hydrogen. On the one hand we have the spectrum, itself, as shown in the figure which is drawn without reference to scale. The lines fall into four series which have been observed experimentally, and others not indicated, each of which has lines converging to a dotted limiting line. Each line is due to light of a particular wave length λ, and when the figure is drawn to scale the lines



are spaced proportionally to their wave numbers $\nu = 1/\lambda$. It has been found empirically that these wave numbers are given with great accuracy by the famous Balmer formula

$$\nu = R(1/m^2 - 1/n^2),$$

in which R is a constant and m and n are integers. The Lyman series is the sequence of lines corresponding to $m = 1$ and arbitrary positive integral values for n, the Balmer series to

$m=2$, and so on. Associated with this picture was the physicist's conception of a hydrogen atom as a dynamical system consisting of a heavy positively charged nucleus with a much lighter electron carrying an equal negative charge revolving around it. What could be the connection between the two?

To answer this question, Bohr assumed that the electron is not free to move in all of the orbits which would be permitted by ordinary dynamical theory, but only in any one of a denumerable infinity of orbits, the so-called stationary states of motion, selected in a way which will presently be described. Radiation was supposed to be emitted only when the electron fell from one of these orbits to another in which it had a smaller total energy, and the wave number of the radiation emitted was proportional to the difference E_1-E_2 between the total energies of the atomic system in the initial and final states. This somewhat crude dynamical picture corresponds to the fact that hydrogen gas does not emit radiation except when suitably excited, that is, except when the electrons in its atoms have been knocked by a bombardment of some sort into stationary states with larger total energies from which they can fall to states with smaller energy-values. Similar assumptions were made for much more complicated atoms with many electrons revolving around their nuclei.

The calculus of variations was of assistance in the specification of the stationary states and the determination of their orbits. It provided canonical transformations from the variables t, q_k, p_k of the atomic system to new canonical variables t, w_k, J_k which had among other properties those described in equations (8) and the following equations of §3 above. For an atomic system these equations do not involve t and have the somewhat simpler form

$$p_k = S_{q_k}(q, J), \quad w_k = S_{J_k}(q, J), \tag{15}$$

$$H(q, S_q) = E(J). \tag{16}$$

The equations of the trajectories in the new variables are identical with equations (9) which show among other things that the J's are really constants of integration. In accordance with assumptions which had proved successful in the simplest cases the quantum orbits were determined by values of the J's of the special form $J_k=n_kh$, where n_k is an integer and h is Planck's

constant, and the frequency $\nu = c/\lambda$ of the emitted radiation when the system fell from a stationary state of total energy E_1 to one of total energy E_2 was to be given by the formula $\nu h = E_1 - E_2$. This last equation, when the extremals for two stationary states of the hydrogen atom were determined and the total energies E_1, E_2 expressed in terms of the integers m, n belonging to the two states, gave the Balmer formula. It was the first great success of the theory.

But the agreement between theory and experimental results was even more striking than this. When hydrogen gas is excited in a constant electric force field each spectrum line is replaced by a symmetrically situated group of neighboring lines, and a similar effect with a different type of group is produced by a constant magnetic field. By modifying the potential function V of the atomic system to take account of these electric or magnetic forces, formulas for the frequencies ν, agreeing well with those observed on the spectrum, were found by the process described in the preceding paragraphs. It was still more interesting perhaps that the fine structure of the lines of the spectrum, detectable by more delicate measurement, was accounted for by replacing Newtonian mechanics by the mechanics of the relativity theory, as described in §3 above. These were indeed impressive agreements between observation and theory.

For more complicated atoms the mathematics became of course much more difficult and was theoretically somewhat disturbing. Besides the properties listed above the canonical transformation (15) was further required to satisfy conditions of the form

$$J_k = \int p_k dq_k = \int S_{q_k} dq_k, \tag{17}$$

where the integrals were taken throughout a period of the motion of the atomic system. These could be satisfied by transforming the J's to other similar constants when the integral $S(q, J)$ of equation (16) was a sum of functions each containing only one of the variables q_i, and progress had been made in the study of more general cases. But I think it must have been a relief to many when at this stage the quantum theory was given an entirely new form by the advent of the methods of Heisenberg and Schrödinger.

Before we take up the consideration of these later developments we should note that for canonical variables w_k, J_k a consequence of the equations (17) was that each coordinate q of a particle in the periodic atomic system was expressible as a function

$$q = q(w, J) = q(w, nh),$$

which was periodic with period unity in each variable w_k. On account of this periodicity q could be expanded, for a quantum orbit having $J_k = n_k h$, in a Fourier series of the form

$$q = \sum C(n, \tau) e^{2\pi i(\tau_1 w_1 + \cdots + \tau_r w_r)},$$

where the sum is taken for all sets $\tau = (\tau_1, \cdots, \tau_r)$ of positive and negative integers. This is an ordinary multiple Fourier series with the sines and cosines replaced by their values in terms of exponentials, and it is readily provable that $c(n, -\tau)$ and $c(n, \tau)$ are conjugate imaginaries. The corresponding equation for q in terms of the time is found by substituting the values (9) for the w's and has the form

$$q = \sum Q(n, \tau) e^{2\pi i N(n,\tau) t}, \tag{18}$$

where

$$N(n, \tau) = \tau_k E_{J_k}(nh).$$

For a fall from a state defined by integers n to one defined by m's the frequency of the emitted radiation is

$$\begin{aligned} \nu(m, n) &= [E(nh) - E(mh)]/h \\ &= (n_k - m_k) \int_0^1 E_{J_k}[m + \theta(n - m)] d\theta, \end{aligned}$$

by Taylor's formula with integral form of remainder term. This is a mean value between the frequencies $N(n, \tau)$ and $N(m, \tau)$ of the two terms having $\tau_k = n_k - m_k$ in the series (18) for the two orbits. The correspondence principle of Bohr resulted in the assumption that the radiation emitted would be like that of a set of simple oscillators whose q-coordinate has the mathematical representation

$$q(m, n) e^{2\pi i \nu(m,n) t}, \tag{19}$$

where $q(m, n)$ is the corresponding mean value

$$q(m, n) = \int_0^1 Q[m + \theta(n - m), \tau]d\theta, \ (\tau_k = n_k - m_k),$$

between the coefficients $Q(n, \tau)$ and $Q(m, \tau)$. By means of this further rather doubtfully justifiable hypothetical specification of the character of the radiation, theoretical polarizations and intensities of spectrum lines were calculated whose agreements with observed data were close enough to be surprising. This situation is, however, a physical one of rather refined delicacy for a mathematician to discuss, and I will emphasize further therefore only the important result that corresponding to every coordinate q of the atomic dynamical system the theory gave a matrix of elements of the form (19) for which elements situated symmetrically with respect to the principal diagonal are conjugate imaginaries. Such a matrix is called by the mathematicians a Hermitian matrix.

6. *The Theory of Schrödinger.* The quantum theory as proposed by Schrödinger is based upon a very interesting extension of the analogy of Bernoulli between optical paths and mechanical trajectories described near the end of §4 above. His argument can be phrased as follows. It has been seen that the extremals for the integral (13) in the optical field are the same as the mechanical trajectories defined by the action integral (12) when the light-velocity u is given by equation (14). In an optical field, however, the theory of geometrical optics is not sufficient to explain minute optical phenomena, such as interference and diffraction. To have a satisfactory theory of these latter phenomena, it has been found necessary to associate with every field of light rays solutions of the wave equation

$$w_{xx} + w_{yy} + w_{zz} = w_{tt}/u^2. \tag{20}$$

It seems reasonable, therefore, in the effort to explain minute atomic mechanical phenomena, to associate with mechanical trajectories solutions of the wave equation with the value of u defined by (14).

The constant C now assumes a more important role. In order to determine its value, Schrödinger made two assumptions. The first was that the frequency ν of the wave to be associated with the moving particle should be related to the total energy E of the particle by the equation $E = h\nu$, as de Broglie had earlier

suggested; the second was that the group velocity of the one-parameter family of waves associated with the particle for different values of E should be equal for each E to the velocity v of the particle itself. It would take too long to explain here the mathematically somewhat vaguely formulated definition of group velocity. For mathematicians who may be interested the best discussion which I know of is that of Fraenkel.† The result of the two assumptions is that $C=E$, so that the final value for u^2 in the wave equation is

$$u^2 = E^2/[2m(E - V)].$$

If we make the substitution

$$w = \psi(x, y, z)e^{-2\pi i(E/h)t} \tag{21}$$

in equation (20), in order to determine the wave functions which have frequency $\nu=E/h$, we find the new equation for ψ

$$\psi_{xx} + \psi_{yy} + \psi_{zz} + \psi 8m\pi^2(E - V)/h^2 = 0. \tag{22}$$

This is the fundamental equation upon which Schrödinger based his theory of the mechanics of an atom.

With each state of motion of a dynamical system having a potential function $V(x, y, z)$ there is to be associated a real or imaginary solution $\psi(x, y, z)$ of the wave equation (22) which has continuous first and second partial derivatives everywhere in xyz-space, which vanishes at infinity, and which furthermore has first partial derivatives ψ_q with respect to the coordinates $q=x, y, z$ such that the products $r^2\psi_q$ are bounded, where $r^2 = x^2+y^2+z^2$, as is customarily assumed in the potential theory. The determination of solutions ψ with these properties is what the mathematicians call a boundary value problem, the parameter in equation (22) which they would usually denote by λ being E. For the mechanical system of the hydrogen atom, for example, it turns out that the only negative values of E for which the problem has solutions form a denumerable infinity $E_1, E_2, \cdots$, and these are precisely the energies of the Bohr stationary states. Thus the Schrödinger theory singles out automatically the possible energy states for the electron, which in the Bohr theory had to be specified by the special assumptions $J_i=n_ih$.

† *Einführung in die Wellenmechanik*, p. 24.

To every value E_n there corresponds a finite number of solutions

$$w_n = \psi_n e^{-2\pi i(E_n/h)t}$$

of the wave equation (22), and these solutions can be normed and orthogonalized so that

$$\int \psi_n \psi_n^* dxdydz = 1, \quad \int \psi_m \psi_n^* dxdydz = 0, \qquad (m \neq n),$$

where ψ_n^* is the conjugate imaginary of ψ_n and the integrals here and in the following equations are taken over the whole of space. In accord with analogies in electro-magnetic theory it was further assumed that the radiation emitted when the system fell from a state of higher energy E_n to one of lower energy E_m would be that of a simple oscillator with coordinates $q = x, y, z$ of the form

$$\int q w_m w_n^* dxdydz, \tag{23}$$

where w_n^* is the conjugate imaginary of w_n. The expression (23) has the value

$$q(m, n)e^{2\pi i(E_n - E_m)t/h}, \tag{24}$$

with

$$q(m, n) = \int q\, \psi_m \psi_n^* dxdydz.$$

The assumptions of the theory therefore lead automatically to a matrix of elements (24) of the same form as that which arose from the Bohr theory.

Mathematically, the crucial part of the Schrödinger theory is the determination of the characteristic numbers E_n of the boundary value problem associated with the wave equation (22) and their corresponding characteristic functions ψ_n. The methods used are those of separation of variables. Theoretically there is an interesting contact at this point also with the calculus of variations. For the boundary value problem is what the mathematicians call self-adjoint, and every such problem is closely associated with a problem of the calculus of variations. For the simple 3-dimensional case of the Schrödinger theory dis-

cussed above this latter problem is that of finding in the class of functions ψ satisfying the condition

$$\int \psi^2 dxdydz = 1$$

one which minimizes the integral

$$I = \int [\psi_x^2 + \psi_y^2 + \psi_z^2 + \psi^2 8m\pi^2 V/h^2] dxdydz.$$

This is an isoperimetric problem of familiar character whose extremals have the differential equation (22). The relationship between boundary problems and such problems of the calculus of variations has had an elaborate development. It was studied for the special case of minimal surfaces by Schwarz, and for more general cases by Hilbert and many others.

The calculations of the quantum theory of Schrödinger applied to the hydrogen atom and other cases gave theoretical results which for the most part agreed as well or better with observed data than those of the earlier theory of Bohr. Sommerfeld's treatment of the fine structure gave, however, a less satisfactory formula, according to his own statement.† The method used in all cases was to substitute the potential function V of the system in the wave equation, then to determine the energy levels and characteristic functions for the boundary value problem, and finally to analyze the radiation emitted by the oscillators (24). What could not be accomplished directly could sometimes be attained by the use of perturbation methods similar to those used in other mechanical theories.

In closing this section I wish to emphasize again the fact that in the Schrödinger theory, as in the preceding theory of Bohr, the final result was a matrix of elements of the form (24) which characterized the emitted radiation. Furthermore, for solutions of the form (21), the equation (20) or (22) is equivalent to the new equation

$$(25) \quad ihw_t/(2\pi) = -h^2(w_{xx} + w_{yy} + w_{zz})/(8\pi^2 m) + Vw,$$

as one readily sees by eliminating E by means of the equation

$$w_t = -2\pi i E u/h.$$

† *Atombau und Spektrallinien*, Wellenmechanischer Ergänzungsband, pp. 132 ff.

If each momentum p in the Hamiltonian function

$$H(q, p) = (p_x^2 + p_y^2 + p_z^2)/(2m) + V \tag{26}$$

is replaced by the operator $p = -(ih)/(2\pi)\cdot\partial/\partial q$, where q is the corresponding coordinate, and if the resulting operator is designated by $H(q, -(ih)/(2\pi)\cdot\partial/\partial q)$, then equation (26) is expressible in the form

$$\frac{ih}{2\pi}\frac{\partial w}{\partial t} = H\left(q, -\frac{ih}{2\pi}\frac{\partial}{\partial q}\right)w. \tag{27}$$

These remarks have significance for the description in the next section of the more recent forms of the quantum theory.

7. *Present Forms of the Quantum Theory.* The recent forms which the quantum theory has taken are not easy to describe in brief fashion because they involve matrix algebra and mathematical theories of probability as well as much of the mathematics which has been discussed in preceding pages. I should like to try nevertheless to give my fellow mathematicians some idea of the contacts which still remain with the calculus of variations, in spite of the fact that my experience with the newer theories is even more restricted than with the older ones.

The first step toward the newer point of view seems to have been taken by Heisenberg. Because of the artificial character of the mechanical mechanisms involved in the Bohr theory, or perhaps one should say the non-mechanical mechanisms, he proposed to cast those aspects of the theory aside completely and to begin with the matrices (24) or (19) themselves which appear as the prime results of the theory of Schrödinger as well as that of Bohr. To the physicist this seemed a very bold proposal, perhaps because the notions of matrices and the matrix calculus are not a part of his usual mathematical kit. But to the mathematician Heisenberg's proposal would, I think, seem a natural one in view of the results of the preceding theories. The lines of the spectrum have a double or matrix array of frequencies which can be regarded as caused by a similar array of oscillators. What then are the properties of the matrices of oscillators which characterize such spectra? It is this question which would have stumped the mathematician, and it is to Heisenberg's unusual combination of physical experience with

mathematical instincts that an answer has been found which seems in many respects to be successful.

Heisenberg's mathematical postulates are in part simple enough when once they have been formulated. He proposed to replace each coordinate q and momentum p of a classical mechanical system by a Hermitian matrix of the kind described in equation (24), and to replace the energy function $H(q, p)$, expressed as a power series in the p's and q's, by the matrix defined by the same series when the matrices for p and q are substituted. If a suitable interpretation could be found in matrix theory for a Poisson bracket (6) then the differential equations of the system could be expressed in the form

$$\frac{dq_k}{dt} = (q_k, H), \quad \frac{dp_k}{dt} = (p_k, H), \tag{28}$$

analogous to the equations (5).

Heisenberg determined the interpretation of the Poisson bracket for his matrix theory by a limiting process which had been used in the so-called correspondence principle of Bohr and which I somewhat arbitrarily avoided in the preceding pages. The method of Dirac is more accessible to the mathematician. Dirac listed the rules of computation for Poisson brackets as defined for the classical theory and found an equation pointing to the fact that he could make them valid for matrices also by assigning to the bracket (f, g) for two matrices f and g the value $c(fg-gh)$, where c is a constant. The difference $fg-gh$ is not in general zero for matrices. When it is so the matrices are said to be *permutable*. It turns out that the value of c must be $-2\pi i/h$ in order to make the results of the theory agree with experiment.

In classical mechanics the Poisson brackets for the coordinates and momenta have the simple values

$$(q_k, q_l) = 0, \quad (p_k, p_l) = 0, \quad (q_k, p_l) = \delta_{kl},$$

where $\delta_{kk}=1$ and $\delta_{kl}=0$ for $k \neq l$. Hence by analogy, and on account of the definition of the Poisson bracket, the matrices q, p of the quantum theory should satisfy the relations

$$\begin{gathered} q_kq_l - q_lq_k = 0, \quad p_kp_l - p_lp_k = 0, \\ q_kp_l - p_lq_k = ih\delta_{kl}/(2\pi), \\ ih\dot{q}_k/(2\pi) = q_kH - H_{q_k}, \quad ih\dot{p}_k/(2\pi) = p_kH - H_{p_k}, \end{gathered} \tag{29}$$

where the dots indicate time derivatives and the last two equations are the equations (28) with the values of the brackets substituted. For a particular dynamical system the energy matrix $H(q, p)$ is the classical energy function of the system with matrices q, p in place of the classical coordinates and momenta. These are the *fundamental equations* of the quantum theory in matrix form.

A *canonical transformation* is defined by a matrix S whose inverse S^{-1} is the matrix formed from S by replacing all its elements by their conjugate imaginaries and then turning the matrix through 180° on its principal diagonal as axis. In mathematical language S^{-1} is the conjugate transpose of S. If such a matrix is at hand, then the transformation $P=SpS^{-1}$, $Q=SqS^{-1}$ takes every set of matrices p, q satisfying equations (29) into a new set P, Q satisfying equations of the same form. Such transformations were used by Heisenberg to find solutions of equations (29). When, for the case of the hydrogen atom for example, a canonical transformation had been found which took the energy matrix H into a diagonal matrix E, the elements $E(n, n)=E_n$ in the principal diagonal were the energy levels of Bohr, and the frequencies of the elements of the matrices q, p turned out to be the radiation frequencies

$$\nu(m, n) = (E_n - E_m)/h.$$

This matrix calculus of Heisenberg has been given a different form and has been supplemented by a probability interpretation which is based upon analogs of well known algebraic properties of Hermitian matrices. A vector $\psi=(\psi_1, \cdots, \psi_n)$ of complex numbers in n dimensions is said to be *normed* if $\psi_k\psi_k^*=1$, where ψ_k^* is the conjugate imaginary of ψ_k. It is *orthogonal* to a second such vector ψ' if $\psi_k^*\psi_k'=0$. An n-dimensional Hermitian matrix m transforms every vector ψ into another $m\psi$ of the same type, and has a set of n normed and orthogonal vectors $\Psi(m')$, each possessing a real number m' such that $m\Psi(m')=m'\Psi(m')$. The vectors $\Psi(m')$ are called the *characteristic vectors* of the matrix, and the numbers m' are its *characteristic numbers*.† An arbitrarily selected vector ψ is expressible linearly in terms of the characteristic vectors in the form

$$\psi = \sum\Psi(m')c(m'),$$

† They are analogous to the "*eigenfunctions*" and "*eigenvalues*" of Dirac.

where the sum is taken for all the characteristic vectors. If ψ is normed, then it is readily provable that

$$\psi^*\psi = \sum c(m')c^*(m') = \sum |c(m')|^2 = 1,$$

an equation which suggests the possibility of interpreting the real numbers $|c(m')|^2$ as probabilities. These results are stated here for n-dimensional vectors and matrices because it is that case which is familiar to most of us. In the quantum theory the range of the subscript k and of the characteristic values m' may be a continuum, or a combination of a discrete and a continuous range, or of still more general character. For these other ranges the matrix theorems are usually not proved but are assumed to hold by analogy.

In the presentation of the theory as given by Dirac the matrices of Heisenberg are replaced by much more general linear operators which operate on symbols ψ representing the possible states of motion of the dynamical system under consideration. Each state is represented not only by its symbol ψ but also by $c\psi$, where c is an arbitrary complex number, and by a symbol ψ^* to which ψ is related as a vector to its conjugate imaginary. All linear combinations of symbols for states, of the form $c_1\psi_1+c_2\psi_2$, also represent states. For any two states ψ_1 and ψ_2 a scalar product $\psi_1^*\psi_2$ is postulated whose value is a complex number such that $\psi_2^*\psi_1=(\psi_1^*\psi_2)^*$. If this number is zero the two states are said to be *orthogonal*. A symbol for a state may be *normed* so that $\psi^*\psi=1$ by multiplying it by a suitable complex constant. One easily sees that the absolute value of this constant is uniquely determined but its amplitude or phase is arbitrary.

At each time t there is associated with each coordinate of the dynamical system a linear operator q which transforms each state ψ into another and which has properties like those of a Hermitian matrix. The operator q has a system of normed and orthogonalized characteristic states $\Psi(q')$ with characteristic numbers q' such that every state ψ is expressible in the form

$$\psi = \sum_{q'} \Psi(q')c(q'),$$

by analogy with the matrix theory. If the symbol ψ is normed then one finds readily that

$$\sum_{q'} | c(q') |^2 = 1. \tag{30}$$

The fundamental idea of the probability form of the quantum theory is now that when an observation is made at the time t to determine the value of a coordinate of the mechanical system in the state ψ the result is not necessarily unique. The values which may be found for the coordinate are the characteristic numbers q' of the operator q, and the probability that a particular value q' will be found at the time t for the state ψ, is the term $|c(q')|^2$ of the sum (30). The classical case would be the one for which q has a single characteristic number q' and behaves like a Hermitian matrix with all elements in the principal diagonal equal to q' and all other elements equal to zero. Such an operator has but one characteristic number q', and all states ψ are characteristic states, so that a measurement of q always gives the value q' as a certainty.

The momenta of the mechanical system are similarly represented by operators p, and the energy by an operator which is a function $H(q, p)$ of the operators q and p. The equations which characterize these operators for a particular mechanical system are the equations (29) of Heisenberg for reasons similar to those described above in deducing them. It is provable that if two operators p and q satisfy the third of equations (29), then every real number q' is a characteristic number for q, and similarly for p. Moreover, the set of permutable operators $q_1, \cdots, q_r$ has normed and orthogonal characteristic states $\Psi(q') = \Psi(q_1', \cdots, q_r')$ in common, one and only one of which corresponds to every set of real numbers $(q_1', \cdots, q_r')$, and in terms of which every other state is expressible in the form

$$\psi = \int \Psi(q')\psi(q', t)dq'.$$

The integral is taken over the whole of the r-dimensional q'-space and $\psi(q', t)$ is a function of $q_1', \cdots, q_r', t$, which is said to *represent* ψ. The state ψ does not change with the time, but its representative must do so, since the operators q and their characteristic states Ψ are changing. When ψ is normed it follows that

$$1 = \int | \psi(q', t) |^2 dq' = \int \psi^*(q', t)\psi(q', t)dq'.$$

If an observation to determine all of the coordinates q is made at the time t upon a system in the state ψ the possible values for the coordinates are the characteristic sets $(q_1', \cdots, q_r')$, and the probability that a set will be found in the volume element dq' is $|\psi(q', t)|^2 dq'$, where $q' = (q_1', \cdots, q_r')$ is a point in dq'.

It is important and interesting to note that the representation $\psi(q', t)$ of the states which are possible for a given dynamical system are precisely the solutions of Schrödinger's wave equation in the form (27). For it is provable from the equations (29) that the results of operating with q_k, p_k, or H upon a state with representative $\psi(q', t)$ are, respectively, the states represented by

$$q_k'\psi(q', t), \quad -\frac{ih}{2\pi}\frac{\partial}{\partial q_k'}\psi(q', t), \quad \frac{ih}{2\pi}\frac{\partial}{\partial t}\psi(q', t),$$

provided that the phases of the normed characteristic states $\Psi(q')$ are suitably chosen. But since H is also expressible in terms of the q's and p's it follows then readily that

$$\frac{ih}{2\pi}\frac{\partial}{\partial t}\psi(q', t) = H\left(q', -\frac{ih}{2\pi}\frac{\partial}{\partial q'}\right)\psi(q', t),$$

where the operator on the right is formed by replacing in $H(q, p)$ each of q and p by its value as an operator. Thus by reasoning quite different from that of Schrödinger we are led again to the wave equation (27).

A mathematician should perhaps not venture to discuss the physical reasons for things, but I think I owe to my mathematical hearers an effort to explain why the quantum theory has adopted a probability aspect. There are a number of reasons, but I shall speak of only one. The Schrödinger theory provided functions $w(q, t)$ which belonged to the stationary states of an atom but it did not attempt to describe the mechanical character of the atomic system itself except in its identification of the group velocity of a one-parameter family of waves with what would ordinarily be the velocity of a mechanical particle. Such a family or packet of waves may have a resultant amplitude different from zero only in a limited portion of space whose extent may increase or contract as it moves. The coordinates and momenta of such a packet taking the place of the mechanical particle would be to some extent uncertain. In the general

theory this uncertaintly corresponds to the fact that the momenta q are never permutable with the coordinate operators p. Operators have characteristic states in common in general only when they are permutable, and the only states for which the probabilities described above for a particular operator q or p are unity, and the results of observations therefore certain, are the characteristic states of the operator. It follows, therefore, that in general there is no state for which the coordinates and momenta are determinable simultaneously with certainty. The mean square deviations Δq, Δp of a corresponding coordinate and momentum from their mean values can be calculated at least in simpler cases, and it has been shown that $\Delta q \Delta p \geqq h/(2\pi)$, which is the inequality expressing Heisenberg's famous uncertainty principle.

This is as far as we shall have time to go in the discussion of the various modifications of quantum mechanics. In the foundations of the operational or matrix form of the theory the influence of the calculus of variations remains only in the preservation of the notions of canonical equations and transformations, and in the boundary value problems connected with the wave equation of Schrödinger. But other equally interesting mathematical domains have been freely called upon to contribute their quotas. An abstract operational calculus, a kind of general analysis of operations, with an associated theory of probability, seem to be requisites, and the details of the applications of these theoretical mechanisms to the special cases so far studied have also great mathematical interest. This can be said in spite of the fact that at the present time the pure mathematician, in his effort to understand the theory, must plough steadily forward oblivious to a multitude of detailed questions of mathematical accuracy. In view of the success which the theory seems to have had it is not too much to hope that in the end all such questions may be satisfactorily answered or avoided, and it is even possible that the pure mathematician may contribute a valuable mite to this happy result.

THE UNIVERSITY OF CHICAGO

STATISTICAL MECHANICS AND THE SECOND LAW OF THERMODYNAMICS*

BY P. W. BRIDGMAN

One thing that has much impressed me in recent conversations with physicists, particularly those of the younger generation, is the frequency of the conviction that it may be possible some day to construct a machine which shall violate the second law of thermodynamics on a scale large enough to be commercially profitable. This constitutes a striking reversal of the attitude of the founders of thermodynamics, Kelvin and Clausius, who postulated the impossibility of perpetual motion of the second kind as a generalization from the uniformly unsuccessful attempts of the entire human race to realize it. Paradoxically, one very important factor in bringing about this change in attitude is the feeling of better understanding of the second law which the present generation enjoys, and which is largely due to the universal acceptance of the explanation of the second law in statistical terms, for which Gibbs was in so large a degree responsible. Statistical mechanics reduces the second law from a law of ostensibly absolute validity to a statement about high probabilities, leaving open the possibility that once in a great while there may be important violations. Doubtless another most important factor in present scepticism as to the ultimate commercial validity of the second law is the discovery of the importance in many physical phenomena of those fluctuation effects which are demanded by statistical mechanics. It is very hard indeed for one who has witnessed the Brownian motion for the first time to resist the conviction that an ingenious enough engineer might get something useful out of it, and I have no doubt that many in this audience have tried their own hand at designing such a device, and I also have no doubt that their success has been discouragingly nil.

There are other aspects also of the statistical point of view which have become prominent in the last few years, as for ex-

* The Ninth Josiah Willard Gibbs Lecture, delivered at New Orleans, December 29, 1931, under the auspices of this Society, at a joint meeting of this Society, the American Physical Society, and Section A of the A.A.A.S.

ample, the speculations of Eddington on time's arrow and on the meaning of time in general, the speculations of Lewis as to completely reversible time, all the recent concern and new notions about the destiny of the universe as a whole, in which thermodynamic arguments play a most important part, and it is of course notorious that the notions of probability, which are fundamental in statistics, are at the very bottom of wave mechanics.

It is evident, therefore, that the statistical point of view entails consequences important both conceptually and practically. In the hope of helping a little to a better understanding of the situation I propose today to examine a few of the implications and consequences of the statistical point of view. The program is a very modest one, and I hope you will have no expectation of a final solution of any of these difficult questions; my primary purpose is to awaken a more vivid self consciousness of what the situation actually is. I shall be mostly concerned with the classical statistics, and shall have less to say about questions raised by wave mechanics; we shall find that the questions raised by the classical point of view are sufficiently fundamental. I shall throughout adopt the point of view that I have called *operational*, that is, I shall seek the meaning of our statements and concepts by trying to analyze what it is that we do when we are confronted with any concrete physical situation to which we attempt to apply the concept or about which we make the statement.

It is in the first place most important to realize that the statistical method, in which the notions of probability are fundamental, has, when applied to the understanding of physical situations, certain inherent, unique, logical characteristics, so that any account which statistics can give of physical phenomena must have an entirely different aspect from that of such a method of approach as classical mechanics, for example. The reason for this is that probability is not a notion which can be applied to concrete individual events, whereas we demand that we understand, or predict, or control, the individual event. I demand to know what will happen when I throw this particular stone in the air, or explode this particular charge in the cylinder of a gas engine. Ordinary mechanics gives an unequivocal answer, and in general the explanations of ordinary me-

chanics make direct contact with just such concrete individual physical events. But the notions of probability have no such application to individual events, and in fact the notion of probability is meaningless when applied to an individual event. The proof of this is given by mere observation of what we do in applying the notion of probability. Suppose that I show you a die and remark that I intend to throw it in a minute. You volunteer the information that the probability is one-sixth that the throw will be a six. I am sceptical and ask you to justify your statement by pointing out the property of the event, when it takes place, that can be described as a probability of one-sixth for a six. I then make the throw and get a six. What possible characteristics has this single event that can justify your statement? Your statement has immediate meaning only when applied to a long sequence of events, or when applied to the *construction* of the die and the method of throwing it. Even when applied to a sequence of events there is always an unbridgeable logical chasm thwarting a precise application of the notion of probability to any actual sequence. Consider, for example, the classical example of tossing a coin. In practice our first concern is to determine whether the coin is a fair coin, that is, whether it is equally likely that heads or tails appear. Suppose that we make a million throws, and find the excess of one or the other not to be more than 1,000. Then we are likely to say that the coin is fair. But, logically, we are bound to recognize that the coin may have been loaded so that heads were, perhaps, nine times as likely to appear as tails, only that we had happened on one of those excessively rare sequences in which as many tails appeared as heads. Rigorously, there can be no method by which we can be sure that all our past experience has not been one of these excessively rare sequences which logically we are bound to recognize as possible in any statistical assembly.

Passing over these logical difficulties in applying precisely the notions of probability to any actual physical situation, it is evident, I think, that when applied to individual events, probability can have only a secondary or derived meaning. I believe that an examination, as the operational point of view prescribes, of what one does, will show that the meaning to be ascribed to the probability of individual happenings is to be found in the

rules of the *mental* game that one plays in thinking about and trying to understand the individual events. This has important consequences when we attempt to use the notions of probability in building physical theories. We must recognize in the first place that any physical theory demands the construction of some sort of model. Now any model involving notions of probability is of necessity more remote from the physical situation and more esoteric in character than the more usual and naive models, such as are offered by ordinary mechanics. For any statistical model involves conventionalized events to which the notions of probability are by definition applicable, although the notion of probability does not apply at all to the concrete physical events which are the counterpart of the events of the model. It is therefore not surprising that the connection between the properties of the statistical model and the corresponding physical system is somewhat different from the connection in the more ordinary sorts of model.

It would be possible to digress considerably here to discuss the properties which we demand in our models, and the uses to which we put them. The subject is fairly obvious, however, and I believe we can safely assume an understanding of the essential features. The least exacting demand that we make of a model is that it serve as a calculating device, by which we can predict actual physical happenings, and for this purpose any sort of consistent correspondence between the model and the physical happening is satisfactory. The simplest way in which the statistical model can satisfy this simplest and minimum demand is evidently that actual happenings in the physical system shall correspond to high probabilities in the statistical model. This furthermore seems to be the only possibility and all that we can do; in the model there are no certainties, only probabilities, some of which, it is true, may be very close to unity, whereas in the physical system there are invariable happenings, as for example, a cake of ice *always* melts when it is heated above the melting point, or the external atmosphere *always* rushes into an exhausted electric light bulb when it is cracked. On the other hand, it is most natural to say that low probabilities in the model correspond to infrequent occurrences in the physical system. But to go further, and specify exactly how close to unity we shall demand that the probability be that

is to correspond to an invariable happening, is not so easy, and there seems to be a certain unavoidable fuzziness here defeating every endeavor to make sharp connection between the model and actuality. There is a still greater difficulty in giving a precise physical significance to events in the model of low probability; we shall return to this question. For the present the important point for us is that any statistical model is in peculiar degree purely a *paper and pencil* model and peculiar limitations may be expected in the use of such a model.

If the only demand that we put on the model were that it should serve as a computing device, the situation would be comparatively simple and could be quickly dismissed. But as a matter of fact we make the more exacting demand that the model enable us to *understand* the physical situation, and to this end we demand that there be a further correspondence between the properties of the model and of the physical system. Since we do not usually make the extreme demand that the model enable us to understand *all* the physical properties of the system at once, we usually do not demand that there be an exhaustive correspondence between the properties of the physical system and those of the model, but we are satisfied with a correspondence of those properties only which are pertinent for our immediate purposes. Thus for the discussion of the thermodynamic properties of a perfect gas, a model is usually sufficient in which the molecules of the actual gas are replaced by perfectly elastic spheres or ellipsoids, although such a model gives no hint of the optical properties of the actual gas. It is curious, however, that even for thermodynamic purposes we would not be satisfied with a model in which the number of fictitious molecules is not equal to the number of actual molecules which various other sorts of physical evidence lead us to ascribe to the actual gas.

If, now, we are attempting to find a model for the thermodynamic properties of a gas, we see that the accepted models which satisfy these additional requirements go far beyond the original demands, for in such models we encounter all the phenomena of fluctuations. Strictly, such a model never comes to equilibrium, and cannot therefore possess any property which strictly corresponds to temperature as defined classically in terms of equilibrium states. The remarkable fact, of course,

is that the fluctuations of the model were found to correspond to fluctuations in the physical system, as shown by the Brownian motion, and in consequence we have now come to recognize that temperature is physically only an approximate concept, instead of the exact concept originally assumed in thermodynamics. This gives us at once one possible method of dealing with the second law and its apparent violations. If we choose to formulate the second law by the statement that dQ/T is an exact differential, then by its very form it applies only to those situations to which the temperature concept applies, and since the temperature concept never exactly applies to any physical situation, we are left with a law which may be rigorously exact in the limit, but which applies to no actual situation. This method of meeting the situation may perhaps be satisfactory to the pure logician, but to the individual interested in filling his pockets by bootlegging entropy, such considerations will appear as uninteresting and as ineffectively legalistic as the restrictions which the more ordinary sort of bootlegger fails to recognize.

It is probably not possible to set up a mechanistic model of a purely thermodynamic system, for the thermodynamic system knows no details, but only pressure and volume and temperature, whereas it is the essence of a mechanical model that it contain details which can have no counterpart in the physical system in so far as it is purely thermodynamic. This prepares us to recognize that the concepts of thermodynamics have no absolute validity, but are relative to the operations, and in particular to the scale of the operations which we use. Thus a fluid in turbulent motion may have a temperature from the point of view of a thermometer with a bulb several centimeters in diameter, but may have no temperature from the point of view of a minute thermo-couple such as biologists have recently used in probing the interior of single cells. Or again, the entire body of phenomena to which the so-called third law is applicable would immediately appear in a different aspect and the third law would no longer be applicable if the operation of taking atoms apart and recombining them was added to the other permissible operations of a more conventional character. In recent papers Dr. David Watson has discussed some of the consequences of a recognition of some of the other relativistic characteristics of the entropy concept.

There seem to be two diametrically opposite and equally natural reactions to an appreciation of this situation. The first reaction is that of the younger generation, part of which, at least, expects to discover some day in the realm of microscopic operations the possibility of a profitable entropy bootlegging business, and there is, secondly, the reaction of the other school, which is convinced that the second law involves something absolutely fundamental, and that any formulation in terms involving relative magnitudes or permissible operations can be only an incomplete formulation.

In the endeavor of the second school to find a more fundamental formulation it is natural to attempt to capitalize the striking success that the statistical point of view has already had in dealing with the phenomena of fluctuations. One recent and well known attempt in this direction is that of Eddington to appraise the second law in terms of what is essentially a shuffling process. Eddington sees in the universal tendency for entropy to increase, or for a system to pass from a less to a more probable configuration, the analog of what happens when a pack of cards is shuffled. Now although there may be strong points of analogy between these two processes, it seems to me that there are also essential differences, and that the analogy on the whole is not a happy one. Shuffling acquires meaning only when we are able to apprehend the cards as individuals by marking them so that we can identify them as individuals, but in such a way that there shall be absolutely no effect on the shuffling process. The picture that Eddington had in mind was a pack of cards freshly received from the manufacturer, arranged in suits and by rank in the suits, and then shuffled and losing all trace of its original arrangement. But as G. N. Lewis has remarked, the arrangement of the cards in suits is one of entirely arbitrary significance; from the point of view of some other game than whist the initial arrangement was already a completely shuffled arrangement as well as all the subsequent ones. Or we may look at it in another way. Imagine an infinite sequence of deals, the cards being partially shuffled between each deal, and suppose that a complete record is kept of all the hands. Somewhere in this sequence there will be deals in which the cards are distributed among the hands by suit and the arrangement in each hand is by rank. If one examines the record it will be found

that as one proceeds away from the exceptional deal, both forward and backward in the sequence, all trace of the regular arrangement becomes obliterated. That is, shuffling with respect to any configuration selected arbitrarily as of special significance is symmetrical with respect to past and future time, and the analogy with the thermodynamic situation disappears. There are other difficulties with the shuffling point of view. We have seen that shuffling has meaning only when the cards can be identified as individuals. But if the cards can be *handled* as individuals, they can be unshuffled, as any whist player will demonstrate to you within five seconds of picking up his hand. To see in the shuffling process something analogous to the inexorableness of the increase of entropy in nature involves the thesis that, although it may be possible to identify the elements in a physical situation, there is some restriction in nature which prevents us from treating these elements separately and sorting them out. This, it seems to me, is a hard doctrine. There is no suggestion of such a state of affairs in any picture offered by the classical mechanics, and classical mechanics was all that Eddington had in mind. Until the reason is elucidated for this surprising inability of ours to handle what we can see, I believe that the shuffling analogy must be judged only to obscure the situation rather than to clarify it.

At first glance, the possibility of understanding the strange restriction against touching what we can see is even more remote from the wave mechanics point of view than from the classical, for seeing is now to be considered as a kind of touching, namely touching with a photon. It may well be, however, that we have here the key to the ultimate solution, for it is not sufficient merely to touch, but the touch must also control; such a kind of touch doubtless requires the cooperation of a great many photons, and thus will be less possible of attainment than the touch by a single photon which is sufficient for recognition. But this point of view we do not follow further here.

Not only does Eddington see in shuffling the complete analogy of the inexorable increase of entropy, but he sees something even more fundamental, namely an explanation of the properties of time itself, and in particular why it is that time is unsymmetrical and flows only forward. This he has expressed by saying that the increase of entropy of the universe is what it is that gives direc-

tion to time's arrow. This conception deals with such fundamental matters and has been hailed in so many quarters as being of such unique profundity, that we may be pardoned for stopping for an examination of what is involved. Eddington sees the crux of the matter in an essential difference between the equations of ordinary mechanics, including electrodynamics, and thermodynamics. The equations of ordinary mechanics and of electrodynamics, which express what Eddington calls primary law, are of such a character that the differential of time may be reversed in sign with no change in the equations, as, for example, in the equation for a falling body, $d^2s/dt^2 = -g$. It is evident enough that the equation is unaltered if the sign of dt is changed; the question is what is the significance of this observation? The significance that Eddington ascribes to it is that the equation is unaffected by a reversal of the direction of flow of time, which would mean that the corresponding physical occurrence is the same whether time flows forward or backward, and his thesis is that in general there is nothing in ordinary mechanical occurrences to indicate whether time is flowing forward or backward. In thermodynamic systems, on the other hand, in which entropy increases with time, time enters the differential equation as the first derivative, so that the direction of flow of time cannot be changed without changing the equation. This is taken to indicate that in a thermodynamic system time must flow forward, while it might flow backward in a mechanical system.

As thus expressed there seems to be considerable vagueness about some of the ideas; this vagueness I find in Eddington's original formulation. Such vagueness is natural and perhaps inevitable in a popular exposition, but if we are to understand from the operational point of view what is involved here, we must try to be more precise. Careful analysis has not yielded to me more than the following as an exact statement of what is involved. Imagine a closed system, and an assistant with a note book in which at a given instant of time he notes all the data necessary to characterize completely the configuration of the system. At a later instant of time he records the corresponding data in another note book. He then gives us the note books and we try from a study of them to determine which set of data was recorded at the earlier instant of time. If the system was a

thermodynamic system we can make the decision, because the entropy increases with time, but if the system was a mechanical system we cannot decide which note book was used first, because examples could be set up for either one or the other. Eddington expresses this difference by saying that the direction of flow of time has no significance in the mechanical system.

It is, of course, true that the differential equation of the mechanical system is differently constructed from that of the thermodynamic system, but the significance of the difference does not need to be formulated as does Eddington. We must in the first place remember that the equation of the mechanical system, for example, $d^2s/dt^2 = -g$, applies not only to a single system, that is to a single falling body, but applies as well to a family of systems. The equation has the property that, corresponding to every specific system, with its particle at a definite point moving with a definite velocity at a definite instant of time, another system is possible with its particle in the same position at the same instant, but with a negative velocity. This is because the equation is of the second order, and gives on integration two constants, which may be so adjusted as to give any position and any velocity at any instant of time. Imagine the second system set up; as time goes on it will trace out in reverse sequence the positions of the first system, as may be seen from the equation itself, which may be written

$$\frac{d^2s}{dt^2} = \frac{dv}{dt} = \frac{d(-v)}{d(-t)} = -g.$$

It is this fact which makes it impossible to decide, in our example above, which note book was used first, because there is no way of telling from the entries alone whether they applied to the first or the second system. But in no case is there any question of time flowing backward, and in fact the concept of a backward flow of time seems absolutely meaningless. For how would one go to work in any concrete case to decide whether time were flowing forward or backward? If it were found that the entropy of the universe were decreasing, would one say that time was flowing backward, or would one say that it was a law of nature that entropy decreases with time? It seems to me that in any operational view of the meaning of natural concepts

the notion of time must be used as a primitive concept, which cannot be analyzed, and which can only be accepted, so that it is meaningless to speak of a reversal of the direction of time. I see no way of formulating the underlying operations without assuming as understood the notion of earlier or later in time.

Lewis in a recent paper *The symmetry of time in physics* adopts a point of view in some respects similar to that of Eddington, although Lewis would certainly disclaim more than very partial agreement with Eddington. Lewis speculates about the fundamental significance of the symmetry of time in physics. His point of view takes its origin from the four-dimensional representation of events employed in relativity theory, and the consequent reduction of all propositions in kinematics to propositions in four-dimensional geometry, rather than from the form of the equations of mechanics, as in Eddington's theory. But both neglect what I believe to be the most important aspect of the situation. Both the equations giving the motion of the system and the four-dimensional representation of the motion are only a small part of the story. The equations are without significance unless the physical operations are also defined by which numerical values are assigned to the various symbols of the equations. For instance, in treating a falling body, we need in addition to the equation itself a set of directions for the use of the equation, in which it is set forth, among other things, that s is the number obtained by making with a meter stick certain manipulations connected with an arbitrary origin and the instantaneous position of the falling body. Similarly in the four-dimensional representation, we must know the physical operations by which the numerical values of the coordinates are obtained which go into the four-dimensional diagram. If one examines the operations by which meaning is given to the symbols which occur in the equations or to the coordinates in the geometrical representation it will be found that the time concept has to be assumed as primitive and unanalyzable, for the operations essentially assume that the operator understands the meaning of later and earlier in time. For example, in order to find the velocity of a particle, one has to observe its position at some one instant of time and combine with this in a prescribed way the result of another observation at a later instant. If one does not intuitively understand what is meant by a later in-

stant, there is no method of formulating the operations. The same situation is involved in specifying a thermodynamic system. One of the variables is the temperature; it is not sufficient merely to read at a given instant of time an instrument called a thermometer, but there are various precautions to be observed in the use of a thermometer, the most important of which is that one must be sure that the thermometer has come to equilibrium with its surroundings and so records the true temperature. In order to establish this, one has to observe how the readings of the thermometer change as time increases.

This point of view, that the schedule of operations by which the symbols acquire meaning is as important a part of the physical situation as the relations which are found to hold between the symbols themselves, has an important bearing on a very widely spread tendency in modern physics and science in general to see nothing as significant except the relations, and so to reduce all science to a kind of topology. It is this point of view that is at the bottom of Einstein's philosophy when he says, for example, that all that is observed is a series of space-time coincidences, and which Eddington expresses by saying that nature may be reduced to a series of pointer readings. If one grants that the ultimate object of physics is to establish a certain sort of relation between us and the world of our senses, to speak with a certain monstrous naïvété, I do not see how it is possible to discard as irrevelant the fact, for example, that the fourth coordinate in the four-dimensional geometry of relativity has to be obtained by an entirely different sort of operation from the other three coordinates, or to regard the entire situation as exhaustively characterized by the relations between the numbers, irrespective of how they are obtained.

We return now to a further consideration of our statistical model and the methods by which we shall establish a correspondence between its properties and those of the physical system. Hitherto we have been gratifyingly successful; events in the model of overwhelmingly high probability correspond to invariable happenings in the physical system, and less common events in the model, such as fluctuation phenomena, are found to be prophetic of a previously unsuspected new domain of physical effects, typified by the Brownian motion. Encouraged by this success, it is natural to think that we have got hold of

something real, whatever that may mean, and to push our scheme of correlation to the logical limit, and say that all the excessively rare events corresponding to low probabilities in the statistical model are correlated with corresponding rare events in the physical system. Now it is a consequence of the fundamental assumptions which have gone into the usual statistical model, namely that all elementary configurations are entirely independent of each other, so that the probability of any configuration is to be calculated by purely combinatorial methods from the relative number of ways in which the configuration can be realized, that there is some chance of the occurrence of any configuration, no matter how unusual its properties. This would mean that in the corresponding physical system any configuration whatever, compatible with the fixed conditions, would occur occasionally, as, for example, the gas in a box will occasionally automatically all collect itself into one end. This conclusion is indeed taken literally by many experts in statistical mechanics, and in the literature statements are not uncommon, such, for example, as that of Bertrand Russell in a recent magazine article that if we put a pail of water on the fire and watch it for an indefinite time, we shall eventually be rewarded by seeing it freeze. It seems to me that there are a couple of objections that can be made to the conventional treatment of rare occurrences, which I shall now examine.

The first difficulty is with the technical method of calculating the chances of observing a rare configuration, and is concerned only with the model itself, and not with the physical application of the results of the calculations. In computing the chance of any configuration, it is always assumed that the elements of the statistical model are without influence on each other, so that the chance of a given configuration is given by merely enumerating the number of complexions corresponding to the given configuration. For example, in the kinetic theory of gases it is assumed that the location of any molecule and its velocity is, except for the restriction on the total energy and the total volume, independent of the location or the velocity of any or all of the other molecules. It may be proper enough to postulate this for the model, but we know that it cannot rigorously correspond to the physical system, for the molecules of a gas do interact with each other, as shown by the mere fact that they conserve

their total energy, and the transmission of energy from one molecule to another takes place only at a finite rate, so that if, for example, at one instant all the velocity were in a single molecule, we would find that immediately afterward only molecules in the immediate vicinity had any velocity. This means that the assumption of complete independence must be recognized to be only an approximation, and some way of handling this approximation must be devised. The method usually adopted is to cast the problem in the form of inquiring how many observations must be made in order that the chance of observing the desired rare configuration may be one half, for example, choosing the time between observations so long that at each observation all appreciable trace of the previous configuration shall have been obliterated, so that the assumption of independence may apply. The point now is this: the time that one has to wait for the probable obliteration of all traces of a previous configuration becomes longer the rarer the previous configuration; obviously it takes longer for a gas to efface all trace of having been all concentrated in one half of its available volume than to efface the traces of a small local concentration. The situation is, therefore, that not only must we make an increasingly large number of observations in order to hope to witness a rare configuration, but the interval between our observations must also get longer. It is merely the first factor which is usually considered; when both factors are considered it is not at all obvious that the process is even convergent. This point should be subjected to further examination.

There is another difficulty connected with the mere calculation of the probability of rare occurrences presented by quantum theory. All classical calculations assume that the molecules have identity. But the uncertainty principle sets a limit to the physical meaning of identity. It is not possible to observe the position and velocity of any molecule with unlimited precision, but there is a mutual restriction. After an observation has been made, the domain of uncertainty in which the molecule is located expands as time goes on. If the domains of uncertainty of two molecules overlap, then the identity of the molecules is lost, and a subsequent observation will not be competent to decide which molecule is which. The only way of maintaining the identity of the molecules is by making observations at

intervals so frequent that the domains of uncertainty have not had time to overlap. But this time is obviously much shorter than the time between observations demanded by the requirement that all trace of the previous configuration shall have been wiped out. Furthermore, the act of observation, by which the concept of identity acquires meaning, alters in an uncontrollable and unpredictable manner the motion of the molecules, whereas the statistical treatment requires that the molecules be undisturbed between successive observations. It seems, therefore, that the physical properties of actual molecules as suggested by quantum theory are different from those of the molecules of the model, and this would seem to demand at least designing new methods of calculating the chances of rare occurrences.

Apart from these objections, which may be met by the discovery of new theoretical methods of attack, it seems to me that the most serious difficulty with this question of rare states is met in the process of transferring to any actual physical system conclusions based on a study of the corresponding model. Suppose, for example, that we are discussing the problem of the tossing of some particular coin. If the coin is a fair coin, that is, if the chances of heads and tails are even, then our statistical model consists merely of a sequence of one or the other of two events, each of which is as likely to occur at any time as the other, absolutely independently of what may have happened elsewhere in the sequence. The theoretical discussion of this model is very easy, and we are convinced that conclusions drawn from a discussion of the model will apply to the tossing of the coin, always provided that the coin is a fair coin. As a particular problem we may consider the chance of throwing heads ten consecutive times. The chance of this is $(1/2)^{10}$, or $1/1024$, which means that in every 1,000 consecutive throws the chances will be roughly even that there will be somewhere a sequence of 10 heads.* But 1,000 throws are a good many, and it may be that we have never made so many throws, and are content merely to make the prediction that if some one else should make so many throws it would be found to be as we say. But suppose that some one questions the fairness of the coin,

* I am much indebted to Mr. H. M. James for a rigorous solution of the interesting problem involved here. He finds that between 1422 and 1423 throws are necessary for a 0.5 chance of ten or more consecutive heads.

and says that he has reason to think that there is a bias of 10% in favor of throwing tails, so that the chance of a head at a single throw is only 0.45 instead of 0.50. We find now on making the calculation that we shall have to make roughly 10,000 throws in order to have an even chance of getting a sequence of 10 heads; and, in general, that slight imperfections in the fairness of the coin make very large differences in the chance of rare occurrences. In view of this, we feel that it behooves us to make some objective test of the fairness of the coin before we venture to publish our prediction that we are likely to get a sequence of 10 heads in 1,000 throws. We make the most direct test possible by appealing to the fundamental definition of fairness, which is that in a large number of throws the ratio of the number of heads to tails tends to equality. But how many throws are necessary to establish such an equality with satisfactory assurance! There is another theorem here, namely that in n throws the chances are even that we shall have an excess either of heads over tails or of tails over heads of $0.6745\ n^{1/2}$. Neglecting the numerical factor for our rough purposes, this means that if we make a hundred throws the chances are nearly even that the number of heads is somewhere between 46 and 54. To establish the fairness of the coin we would have to make a considerable number of 100 throws at a time and observe whether or not the number of heads clusters between 45 and 54. If, on the other hand, there is a 10% bias in favor of tails, the number of heads will cluster between 40 and 50. The precise number of sequences of 100 throws at a time necessary to convince us that there is no 10% bias in favor of tails obeys no definite criterion, but it is certainly of the order of ten or more, which makes 1,000 or more throws altogether. But this was the number of throws necessary to obtain one of the rare sequences of 10 heads.

The conclusion from all this is plain; in order to establish with sufficient probability that the actual physical system has those properties which are assumed in estimating the frequency of rare occurrences it is necessary to make a number of observations so great that the probability is good that the rare occurrence has already been observed. In other words, purely logical statistical considerations never can justify us in predicting events so rare that they have never yet been observed. A pail of water has never been observed to freeze on the fire; statistical

considerations give us no warrant whatever for expecting that it ever will. Such predictions can be made only on the basis of considerations other than statistical. Thus in the case of the coin, an exact measurement of its dimensions and of the degree of homogeneity of its metal might convince us that the chances of heads and tails were even, because of our knowledge of the laws of mechanics. But when we come to the molecules of a gas or the elements of other physical systems to which the statistical method of treatment is usually applied, we see that there is no method of independently handling the elements, so that the statistical method of testing the validity of our assumptions is the only possible one. This is a most natural situation, because if we were capable of dealing with the elements of the physical system as individuals we could apply more powerful methods than the statistical. Incidentally we may remark how very insensitive the statistical method is in studying elemental properties; this is shown by the example of the coin above, where we had to make something of the order of 1,000 throws to establish an asymmetry of 10%. In many cases, however, the statistical method is doubtless the ultimate and the only method.

Another of the applications of statistical ideas in which there has always been much interest, and especially lately, is to the problem of the ultimate fate of the universe. It is a very common opinion that the second law, in its original classical form, demands the ultimate *heat death* of the universe, because of the inexorable increase of entropy to a final maximum, when all temperature differences shall have been wiped out. The chief mechanism in the ultimate equalization of temperature is obviously the radiation that is continually emitted by the stars. The human mind has, however, shown itself curiously unwilling to accept the prospect of a heat death, and there have been a number of attempts to avoid such an unwelcome conclusion. At least two of these, somewhat similar to each other, are based on the statistical interpretation of the second law. The first of these utilizes the theorem that in a closed mechanical system any configuration, once experienced, is bound to recur after the lapse of sufficient time. According to this view, the universe endlessly goes through cycles of repetition, the so-called Poincaré-Zermelo cycle, of prodigious but calculable duration. The obvious objection to this picture is that in order to realize a

Poincaré-Zermelo cycle the laws of classical mechanics would have to be satisfied with an exactness quite fantastic, hopelessly beyond the possibility of direct or indirect verification. The second attempt to make statistics avoid the heat death rests on the theorem that a statistical assembly is never in complete equilibrium, but is always subject to fluctuations, and these fluctuations may attain any intensity if we only wait long enough. The present state of the universe is then to be regarded only as a fluctuation, with the possibility that similar fluctuations may recur in the future. The difficulty with this point of view is the excessive rareness of the sort of fluctuation corresponding to the present state of the universe compared with the approximate dead level of the *heat death*. The previous considerations apply; this is one of those configurations so rare that one has no right to predict its occurrence unless it has been previously observed. One might predict from purely statistical considerations the occurrence of such a fluctuation in the future if one were sure that one were observing such a fluctuation now. But where is the evidence for this? According to the astronomers the fluctuation has been taking the last 10^{16}, or perhaps now 10^{10}, years or so to smooth itself out to its still considerable roughness, and there is certainly no evidence that before 10^{16} years ago the entropy was decreasing instead of increasing.

There are other objections to an application of the second law to the entire universe. The original formulation of the second law was, of course, restricted to isolated systems. By what logical right can the argument be extended to the entire universe? A natural reply is that relativity theory seems to demand that the universe is finite, so that the whole universe becomes the sort of isolated system demanded by the classical formulation. But I believe that examination will nevertheless show a very important difference betweeen the smaller and the all-embracing closed system. Statistical mechanics, if it is to avoid the difficulties already discussed when applied to any individual physical situation, must make the assumption of molecular chaos. But what in the physical situation gives rise to molecular chaos? If we imagine a gas, for example, in a perfectly reflecting enclosure, and suppose that the molecules are perfectly elastic spheres, then, according to the classical picture, every collision takes place under perfectly definite conditions, so that a mathe-

matician of sufficient power could compute backward from the present configuration to the configuration at any past time, as, for example, when a partition might have been removed from the middle of the compartment. This sort of condition certainly cannot be described as molecular chaos. But the walls are molecular in structure, so that the reflection of the individual gas molecules follows no definite rule. If we regard the molecules in the wall as part of the external universe, and if there is no coordination between the motion of the molecules of the wall and what is taking place in the gas because of the enormous magnitude of the external universe compared with the gas inside, then a physical reason justifying the assumption of molecular chaos is at once apparent. When the entire universe is considered there can be no such justification as this for assuming molecular chaos, but the whole course of events must, from the classical point of view, run a rigorously deterministic course, to which statistical considerations do not apply. It may be objected to this argument that a gas in contact with its walls in the way described above is not an isolated system. It is, of course, not completely isolated, but it is nevertheless as far as the thermodynamic requirements go, which are concerned only with transfers of energy and of heat from the outside. Complete isolation would seem to be incompatible with molecular chaos.

It was intimated at the beginning of this discussion that the *heat death* is supposed to be a consequence of the second law, and that the continuous enormous radiation into empty space of the stars is the most striking manifestation of this tendency. This point of view sees in the emission of every photon by a star part of the inexorable increase of entropy. I believe, however, that this is fallacious, that the relations are different, and that the heat death with which we are confronted as a consequence of continued radiation is an affair of the first law, not of the second, and will take place when all the energy of the universe has been radiated away, not merely when the energy is uniformly distributed. Elementary considerations justify this contention. Consider a body radiating into empty space. It is continually dropping in temperature and losing in energy. The emission of radiation therefore decreases the entropy of the radiating body. Consider next a body in thermal equilibrium with its surroundings; since it is in equilibrium its entropy is constant,

and furthermore it absorbs as much radiation as it emits. Absorption of radiation, therefore, increases the entropy of the absorbing body. This is sufficient to give a straightforward account of entropy changes in radiation problems. Consider two bodies confronting each other in a cavity at constant temperature. A photon leaves one of the bodies, decreasing the entropy of that body. During the passage of the photon across the space separating the two bodies the entropy is to be thought of as associated with the photon and residing somewhere in the intermediate space. When the second body absorbs the radiation, the photon with its entropy disappears from empty space and increases the entropy of the absorbing body. At all stages of the process the entropy of the entire system is unchanged. But now suppose that one body radiates to another at lower temperature. This process is irreversible and is accompanied by an increase of entropy. The first two steps of the process, emission of the photon and passage across the intermediate space, are the same as before, and are therefore accompanied by no net change of entropy. It is only the last stage of the process, absorption by the body at lower temperature, that can give the uncompensated increase of entropy. The mechanism of this increase is to be found in a diffusion of the energy of the photon into the greater number of degrees of freedom corresponding to the lower temperature. Actually, this argument is over-simplified, and must be modified by a consideration of the distribution of the photons through the spectrum, but the details of this point of view need not concern us further. The immediate point for us is that emission of radiation into empty space is not an entropy changing process; the increase of entropy can occur, if it occurs at all, only during the act of absorption.

But what physical evidence have we of the absorption of the light radiated by the stars? To save the situation we must postulate absorption under completely unknown conditions. But is the assumption of such unknown conditions in regions so terribly far beyond access to us in order to save the second law any easier as an intellectual feat than the assumption of other unknown conditions which would defeat the second law? Are we not completely in the dark here, and had we not better admit it?

Finally, I briefly summarize what I believe to be the principal results of this analysis and indicate the possible lines of future

progress. The most important result will be, I hope, a keen realization that in using statistics we are only using a paper and pencil model, which has logical difficulties within itself and difficulties of application to concrete physical situations which are very much greater than the corresponding difficulties with more ordinary sorts of model. Some of these difficulties I believe can never be surmounted, so that the statistical model can never be satisfactorily used by extrapolation either into remote epochs of time, to predict rare events, or into remote reaches of space, to give us an idea of the course of universal evolution. Our model has not given us a satisfactory answer to our initial question as to the possibility of commercially profitable violations of the second law. The answer to this question will probably be found when the wave mechanics point of view has been completely worked out. Some of the other logical difficulties of the classical statistics, I believe, will also be surmounted by the adoption of the wave mechanics point of view, which assumes probability to be a primitive property of the elements of the model, rather than an emergent property resulting from the cooperation of great numbers. Before, however, the wave mechanics thesis of the primitive character of the notion of probability can be accepted, much more experimental work is necessary. If the thesis is supported, as seems probable from the evidence now at hand, this will constitute to a certain extent a defeat of the purpose of the classical statistics, which was to *explain* why many physical assemblages of large numbers of elements obey the rules of probability. But even granted that the primitive character of the notions of probability acquires an experimental verification, it seems to me that some of the logical difficulties will persist, justifying a doubt as to the possibility of ever setting up a logically completely satisfying correspondence between our models and our experience.

HARVARD UNIVERSITY,
RESEARCH LABORATORY OF PHYSICS

NOTE ON DIOPHANTINE AUTOMORPHISMS*

BY LEONARD CARLITZ†

1. *Introduction.* Professor E. T. Bell‡ defines a diophantine automorphism of a form $f \equiv f(x_1, \cdots, x_n)$, where f is a homogeneous polynomial, to be a transformation of f into f^h, h an integer >1, by means of a substitution $X_i = X_i(x_1, \cdots, x_n)$, $(i = 1, \cdots, n)$, so that

$$f(X_1, \cdots, X_n) = f^h(x_1, \cdots, x_n). \tag{1}$$

The first non-trivial instance of an algebraic phenomenon of this kind is due to Eisenstein,§ who noticed that the discriminant of the general binary cubic possesses such an automorphism. Indeed, if we set

$$f(x_1, x_2, x_3, x_4) = x_1^2 x_4^2 - 3x_2^2 x_3^2 + 4x_1 x_3^3 + 4x_2^3 x_4 - 6x_1 x_2 x_3 x_4,$$

$$X_1 = 3x_1 x_2 x_3 - x_1^2 x_4 - 2x_2^3 = -\frac{1}{2}\frac{\partial f}{\partial x_4},$$

$$X_2 = 2x_1 x_3^2 - x_1 x_2 x_4 - x_2^2 x_3 = \frac{1}{6}\frac{\partial f}{\partial x_3},$$

$$X_3 = x_1 x_3 x_4 - 2x_2^2 x_4 + x_2 x_3^2 = -\frac{1}{6}\frac{\partial f}{\partial x_2},$$

$$X_4 = x_1 x_4^2 - 3x_2 x_3 x_4 + 2x_3^3 = \frac{1}{2}\frac{\partial f}{\partial x_1},$$

it may be verified that $f(X_1, X_2, X_3, X_4) = f^3(x_1, x_2, x_3, x_4)$. This follows at once if it be noticed that $f(X)$ is the discriminant of the cubicovariant of the basic binary cubic, and further that all the invariants of the binary cubic are expressible in terms of the discriminant.

Cayley has given a different proof and at the same time a generalization‖ of Eisenstein's result. If we set

---

* Presented to the Society, December 30, 1930.

† National Research Fellow.

‡ This Bulletin, vol. 33 (1927), pp. 71–80.

§ Journal für Mathematik, vol. 27 (1844), p. 105.

‖ A. Cayley, Collected Mathematical Papers, vol. 1, 1889, No. 15, *Note sur deux formules données par MM. Eisenstein et Hesse*, pp. 113–116.

$$u(x) = x_1^2x_8^2 + x_2^2x_7^2 + x_3^2x_6^2 + x_4^2x_5^2 + 4x_1x_4x_6x_7 + 4x_2x_3x_5x_8 - 2x_1x_2x_7x_8 - 2x_1x_3x_6x_8 - 2x_1x_4x_5x_8 - 2x_2x_3x_6x_7 - 2x_2x_4x_5x_7 - 2x_3x_4x_5x_6,$$

$$X_i = \frac{1}{2}\frac{\partial u}{\partial x_i}, \qquad (i = 1, \cdots, 8),$$

then Cayley's result is $u(X) = u^3(x)$. In another paper,* Cayley proves that the Hessian of Eisenstein's quaternary quartic, that is, $f(x_1, x_2, x_3, x_4)$ above, is the product of a numerical constant into the square of f. However, no connection between this property and the automorphic property is indicated.

Another instance of (1) is furnished by the general symmetric determinant of order n: its adjoint is also symmetric and is equal to the $(n-1)$ st power of the original determinant. This is of course true of the general determinant. However the latter is quite trivial since the general determinant, regarded as a function of n^2 independent variables, is *composable*, and clearly all composable forms have the automorphic property (1). It should be noticed that here again the automorphic transformation is expressed in terms of partial derivatives of the first order.

In this note we consider diophantine automorphisms characterized by this property. We assume a form f of degree k in n variables such that

$$(2) \qquad \begin{cases} f(X_1, \cdots, X_n) = \alpha f^{k-1}(x_1, \cdots, x_n), \\ X_i = \sum_{j=1}^{n} c_{ij}\frac{\partial f(x)}{\partial x_j}, \qquad \left|\frac{\partial X_i}{\partial x_j}\right| \not\equiv 0, \end{cases}$$

where α and c_{ij} are numerical. We prove two results.

(I) The transformation defined by the second of (2) is a Cremona transformation, and (but for a linear transformation) is of period two:

$$(3) \qquad X_i(X) = \alpha f^{k-2}\sum_{j=1}^{n}\gamma_{ij}x_j.$$

This result is a sort of converse to a result of L. Weisner,† who shows that from a Cremona transformation of finite period a diophantine automorph may be deduced.

* Loc. cit., No. 54, *Note sur les hyperdéterminants*, pp. 352–355.

† This Bulletin, vol. 33 (1927), pp. 707–712.

(II) If $f(x_1, \cdots, x_n)$ is irreducible (in the field of rationals, say), then the Hessian of f is a constant multiplied into a power of f:

$$H[f] = \beta f^{n(k-2)/k}. \tag{4}$$

This second result shows the connection between the Cayley identities quoted above. From (4) we have as an immediate corollary that for a function satisfying (2) the degree, k, is a divisor of $2n$, n being the number of variables.

2. *The Cremona Property.* Let us write

$$f_j = \frac{\partial f(x)}{\partial x_j}, \quad Y_i(x) = \sum_{j=1}^{n} c_{ji} f_j.$$

Then since

$$\frac{\partial}{\partial f_i} = \sum_{j=1}^{n} \frac{\partial X_j}{\partial f_i} \frac{\partial}{\partial X_j} = \sum_{j=1}^{n} c_{ji} \frac{\partial}{\partial X_j},$$

we find that

$$\begin{aligned} Y_i(X) &= \sum_j c_{ji} \frac{\partial f(X)}{\partial X_j} = \frac{\partial f(X)}{\partial f_i} = \alpha \frac{\partial}{\partial f_i} f^{k-1}(x) \\ &= \alpha \sum \frac{\partial f^{k-1}}{\partial x_j} \frac{\partial x_j}{\partial f_i} = \alpha(k-1) f^{k-2} \sum f_j \frac{\partial x_j}{\partial f_i}. \end{aligned} \tag{5}$$

But

$$\sum x_j f_j = kf,$$

so that

$$x_i + \sum f_j \frac{\partial x_j}{\partial f_i} = k \frac{\partial f}{\partial f_i} = k \sum \frac{\partial f}{\partial x_j} \frac{\partial x_j}{\partial f_i} = k \sum f_j \frac{\partial x_j}{\partial f_i},$$

and

$$x_i = (k-1) \sum f_j \frac{\partial x_j}{\partial f_i}.$$

Substituting into (5), we find

$$Y_i(X) = \alpha f^{k-2} \cdot x_i,$$

and

$$X_i(X) = \alpha f^{k-2} \sum \gamma_{ij} x_j \tag{3}$$

where $\|\gamma_{ij}\| = \|c_{ij}\|^{-1} \cdot \|c_{ij}\|$.

3. *Proof of* (4). It is evident that

$$\left|\frac{\partial Y(X)}{\partial X}\right| \cdot \left|\frac{\partial X}{\partial x}\right| = \left|\frac{\partial Y(X)}{\partial x}\right|.$$

From this fact, and from the equations

$$\left|\frac{\partial Y(X)}{\partial X}\right| = |c_{ji}| \cdot \left|\frac{\partial^2 f(X)}{\partial X_i \partial X_j}\right| = cH[f(X)],$$

$$\left|\frac{\partial X}{\partial x}\right| = |c_{ij}| \cdot \left|\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right| = cH[f(x)],$$

where H denotes the Hessian constructed with respect to the variables indicated, we find

$$c^2 H[f(X)]H[f(x)] = \left|\frac{\partial Y(X)}{\partial x}\right| = \alpha^n \left|\frac{\partial (x_i f^{k-2})}{\partial x_j}\right|.$$

The evaluation of this determinant presents no difficulty. Let us use the notation $\phi_j = \partial\phi/\partial x_j$, and let δ_{ij} denote the Kronecker delta. Then we may write

$$\left|\frac{\partial (x_i\phi)}{\partial x_j}\right| = |x_i\phi_j + \phi\delta_{ij}|$$

$$= -\begin{vmatrix} x_1\phi_1 + \phi & x_1\phi_2 & \cdots & x_1\phi_n & x_1 \\ x_2\phi_1 & x_2\phi_2 + \phi & \cdots & x_2\phi_n & x_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_n\phi_1 & x_n\phi_2 & \cdots & x_n\phi_n + \phi & x_n \\ 0 & 0 & \cdots & 0 & -1 \end{vmatrix}$$

$$= -\begin{vmatrix} \phi & 0 & \cdots & 0 & x_1 \\ 0 & \phi & \cdots & 0 & x_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \phi & x_n \\ \phi_1 & \phi_2 & \cdots & \phi_n & -1 \end{vmatrix} = -\{1, 2, \cdots, n\}, \text{ say.}$$

Then we shall have

$$\begin{aligned}\{1, 2, \cdots, n\} &= \phi\{2, \cdots, n\} - x_1\phi_1\phi^{n-1} \\ &= \phi^2\{3, \cdots, n\} - (x_1\phi_1 + x_2\phi_2)\phi^{n-1} \\ &= -\phi^n - (x_1\phi_1 + \cdots + x_n\phi_n)\phi^{n-1} = -(s+1)\phi^n\end{aligned}$$

if ϕ be homogeneous of degree s. Therefore

$$\left|\frac{\partial(x_i f^{k-2})}{\partial x_j}\right| = (k-1)^2 f^{n(k-2)},$$

and substituting in (6),

$$c^2 H[f(X)]H[f(x)] = \alpha^n (k-1)^2 f^{n(k-2)}.$$

If now we make use of the hypothesis that $f(x_1, \cdots, x_n)$ is irreducible we have immediately

$$H[f(x)] = \beta f^{n(k-2)/k}, \tag{4}$$

the exponent being found by comparing the degrees of both members.

4. *Forms Not Satisfying* (2). It is by no means necessary that a form f having the automorphic property satisfy (2). This is obvious if f be composable and of degree $\neq 3$. Furthermore simple examples exist of forms which are not composable. As an example we remark that

$$(x_1^2 + x_2^2 + x_3^2)^2 = (x_1^2 + x_2^2 - x_3^2)^2 + (2x_1x_3)^2 + (2x_2x_3)^2,$$

but, as is well known, $x_1^2 + x_2^2 + x_3^2$ is not composable.

In another note I shall indicate the construction of a class of forms satisfying (1) but in general not (2); the forms are derived from invariant theory.

CALIFORNIA INSTITUTE OF TECHNOLOGY

ON A MATRIX DIFFERENTIAL OPERATOR*

BY A. K. MITCHELL

1. *Introduction.* H. W. Turnbull† has defined a matrix differential operator and has given several theorems concerning its properties. The first of these theorems is proved by induction, and for the fourth, use is made of a lemma concerning the principal minors of a determinant for which he gives a long and intricate proof. It is the purpose of this note to give a simple and direct proof of this lemma and of the fourth theorem, and to indicate how each of the first three theorems proved by Turnbull may be obtained directly. In the last paragraph will be given, in addition, the results of applying the operator successively to the coefficients of the characteristic equation of a matrix.

2. *The Matrix Differential Operator.* Let

$$E = \begin{Vmatrix} E_1^1 & E_2 & \cdots & E_n^1 \\ E_1^2 & E_2^2 & \cdots & E_n^2 \\ \cdot & \cdot & \cdots & \cdot \\ E_1^n & E_2^n & \cdots & E_n^n \end{Vmatrix}$$

be an n-rowed square matrix whose n^2 elements are treated as independent variables. This matrix is characterized by a typical element, and we shall write $E=E_s^r$, where r denotes the row and s the column. In this way if we have another matrix $F=F_s^r$, we shall write,‡ from the law of multiplication of matrices,

$$EF = E_\alpha^r F_s^\alpha .$$

We observe that

$$EF = E_\alpha^r F_s^\alpha = F_s^\alpha E_\alpha^r ,$$

whereas

$$FE = F_\alpha^r E_s^\alpha = E_s^\alpha F_\alpha^r \neq EF.$$

* Presented to the Society, March 26, 1932.

† H. W. Turnbull, *On differentiating a matrix*, Proceedings of the Edinburgh Mathematical Society, (2), vol. 1, Part 2 (1928), pp. 111–128.

‡ Throughout this paper repeated Greek suffixes will denote summation from 1 to n.

The matrix differential operator Ω is defined as follows:

$$\Omega = \begin{Vmatrix} \partial/\partial E_1^1 & \partial/\partial E_1^2 \cdots \partial/\partial E_1^n \\ \partial/\partial E_2^1 & \partial/\partial E_2^2 \cdots \partial/\partial E_2^n \\ \cdots\cdots\cdots\cdots \\ \partial/\partial E_n^1 & \partial/\partial E_n^2 \cdots \partial/\partial E_n^n \end{Vmatrix}.$$

We note that the order of suffixes is transposed from that of E. The effect of the operator Ω is defined by the ordinary law of multiplication of matrices; thus $\Omega F = \partial F_\sigma^\alpha / \partial E_\tau^\sigma$. For a scalar function f we have $\Omega f = \partial f / \partial E_r^s$.

3. *Proof of Turnbull's Theorems.* Now let I_s denote the sum of all the s-rowed principal minors of the determinant of the matrix E. The fourth theorem given by Turnbull, which we shall prove directly, is then as follows.

THEOREM.

$$\delta_s^r - \Omega I_1 = 0, \quad E - \delta_s^r I_1 + \Omega I_2 = 0,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$E^{n-1} - I_1 E^{n-2} + \cdots + (-1)^{n-1}\delta_s^r I_{n-1} + (-1)^n \Omega I_n = 0.$$

The proof will depend upon the properties of the generalized Kronecker delta* $\delta_{s_1 \cdots s_k}^{r_1 \cdots r_k}$ which, if the superscripts are distinct from each other and the subscripts are the same set of numbers as the superscripts, has the value $+1$ or -1 according as an even or an odd permutation is required to arrange the superscripts in the same order as the subscripts; and which, in all other cases, has the value zero. In particular $\delta_s^r = 0$ if $r \neq s$, $\delta_s^r = 1$ if $r = s$. Then, I_s being the sum of all the s-rowed principal minors of E, we may write:

$$I_1 = \delta_\beta^\alpha E_\alpha^\beta,$$
$$I_2 = \frac{1}{2!}\delta_{\beta_1\beta_2}^{\alpha_1\alpha_2} E_{\alpha_1}^{\beta_1} E_{\alpha_2}^{\beta_2},$$
$$\cdots\cdots\cdots\cdots$$
$$I_s = \frac{1}{s!}\delta_{\beta_1 \cdots \beta_s}^{\alpha_1 \cdots \alpha_s} E_{\alpha_1}^{\beta_1} \cdots E_{\alpha_s}^{\beta_s},$$

* See F. D. Murnaghan, American Mathematical Monthly, vol. 32 (1925), p. 233; and O. Veblen, *Invariants of Quadratic Differential Forms*, p. 3.

from which we have immediately

$$\Omega I_1 = \partial \delta^{\alpha}_{\beta} E^{\beta}_{\alpha} / \partial E^{s}_{r} = \delta^{r}_{s},$$
$$\Omega I_2 = \delta^{r}_{s} I_1 - E^{r}_{s}.$$

Continuing in this way we obtain by an easy calculation*

$$\Omega I_n = \delta^{r}_{s} I_{n-1} - I_{n-2} E^{r}_{s} + I_{n-3} E^{r}_{\alpha_1} E^{\alpha_1}_{s} + \cdots + (-1)^{t+1} I_{n-t} E^{r}_{\alpha_1} \cdots E^{\alpha_{t-2}}_{s} + \cdots + (-1)^{n+1} E^{r}_{\alpha_1} E^{\alpha_1}_{\alpha_2} \cdots E^{\alpha_{n-2}}_{s}.$$

Transposing the left members of these equations we have the theorem stated above.

From the foregoing we see that

$$\partial I_{p-1} / \partial E^{s}_{\alpha} = \delta^{\alpha}_{s} I_{p-2} - I_{p-3} E^{\alpha}_{s} + \cdots + (-1)^{t-1} I_{p-t} E^{\alpha}_{\alpha_1} E^{\alpha_1}_{\alpha_2} \cdots E^{\alpha_{t-2}}_{s} + \cdots + (-1)^{p} E^{\alpha}_{\alpha_1} E^{\alpha_1}_{\alpha_2} \cdots E^{\alpha_{p-3}}_{s}$$

and

$$\partial I_p / \partial E^{s}_{r} = \delta^{r}_{s} I_{p-1} - I_{p-2} E^{r}_{s} + I_{p-3} E^{r}_{\alpha} E^{\alpha}_{s} + \cdots + (-1)^{t-1} I_{p-t} E^{r}_{\alpha_1} \cdots E^{\alpha_{t-2}}_{s} + \cdots + (-1)^{p+1} E^{r}_{\alpha_1} \cdots E^{\alpha_{p-3}}_{s},$$

from which it is evident that

$$\partial I_p / \partial E^{s}_{r} = \delta^{r}_{s} I_{p-1} - (\partial I_{p-1} / \partial E^{s}_{\alpha}) E^{r}_{\alpha}.$$

This is the lemma used by Turnbull.

By induction Turnbull proves the following formula:

$$\Omega E^{r} = n E^{r-1} + s_1 E^{r-2} + \cdots + s_{j-1} E^{r-j} + \cdots + s_{r-1},$$

where s_r is the sum of the rth powers of the latent roots of the matrix E (that is, $s_r = \lambda_1^{r} + \lambda_2^{r} + \cdots + \lambda_n^{r}$).

Now by a theorem due to Sylvester,† the latent roots of any rational function of a matrix E are the corresponding functions of the latent roots of E. In particular the latent roots of E^r are $\lambda_1^{r}, \lambda_2^{r}, \cdots, \lambda_n^{r}$. From

$$\Omega E^{r} = \partial E^{\beta}_{\alpha_1} E^{\alpha_1}_{\alpha_2} \cdots E^{\alpha_{r-1}}_{s} / \partial E^{\beta}_{r},$$

by differentiation, we obtain

---

* See *The derivation of tensors from tensor functions*, American Journal of Mathematics, vol. 53 (1931), p. 198.

†Sylvester, *Mathematical Papers*, vol. 4, p. 133.

$$\Omega E^r = \delta_\beta^\beta \delta_{\alpha_1}^{\ r} E_{\alpha_2}^{\alpha_1} \cdots E_s^{\alpha_{r-1}} + E_{\alpha_1}^{\beta} \delta_\beta^{\alpha_1} \delta_{\alpha_2}^{\ r} E_{\alpha_3}^{\alpha_2} \cdots E_s^{\alpha_{r-1}}$$
$$+ \cdots + E_{\alpha_1}^{\beta} E_{\alpha_2}^{\alpha_1} \cdots E_{\alpha_j}^{\alpha_{j-1}} \delta_\beta^{\alpha_j} \delta_{\alpha_{j+1}}^{\ r} E_{\alpha_{j+2}}^{\alpha_{j+1}} \cdots E_s^{\alpha_{r-1}}$$
$$+ \cdots + E_{\alpha_1}^{\beta} E_{\alpha_2}^{\alpha_1} \cdots E_\beta^{\alpha_{r-2}} \delta_s^{\ r}.$$

But the sum of the latent roots of a matrix is equal to the sum of the elements of the leading diagonal of the matrix, so that we have by Sylvester's theorem, $s_1 = E_\alpha^\alpha$, $s_2 = E_\alpha^\alpha$, $E_\alpha^{\alpha_1}$, $\cdots$, $s^r = E_{\alpha_1}^{\alpha} E_{\alpha_2}^{\alpha_2} \cdots E_\alpha^{\alpha_{r-1}}$. Hence, from the above expression for ΩE^r,

$$\Omega E^r = nE^{r-1} + s_1 E^{r-2} + \cdots + s_{j-1} E^{r-j} + \cdots + s_{r-1}.$$

In like manner each of the Theorems I and III of Turnbull may be proved directly.

4. *Conclusion.* Let us now consider the effect of repeating the application of the operator Ω on the functions $I_1, I_2, \cdots, I_n$ defined above. The I's so defined are the coefficients of the characteristic equation* of the matrix E. From the definition of I_p and of Ω it is readily seen that

$$\Omega I_p = [1/(p-1)!]\delta_{s\beta_2\cdots\beta_p}^{r\alpha_2\cdots\alpha_p} E_{\alpha_2}^{\beta_2} \cdots E_{\alpha_p}^{\beta_p},$$
$$\Omega^2 I_p = [1/(p-1)!](\partial/\partial E_r^\alpha)\delta_{s\beta_2\cdots\beta_p}^{\alpha\alpha_2\cdots\alpha_p} E_{\alpha_2}^{\beta_2} \cdots E_{\alpha_p}^{\beta_p}$$
$$= -[(p-1)/(p-1)!]\delta_{s\alpha\beta_3\cdots\beta_p}^{r\alpha\alpha_3\cdots\alpha_p} E_{\alpha_3}^{\beta_3} \cdots E_{\alpha_p}^{\beta_p}$$
$$= -[(n-p+1)!/(n-p)!]\Omega I_{p-1},$$

and

$$\Omega^2 I_{p-1} = -(n-p+2)\Omega I_{p-2}.$$

Therefore

$$\Omega^3 I_p = [(n-p+2)!/(n-p)!]\Omega I_{p-2},$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$\Omega^p I_p = (-1)^{p+1}[(n-1)!/(n-p)!]\Omega I_1$$
$$= (-1)^{p+1}[(n-1)!/(n-p)!]\delta_s^{\ r}.$$

When $n = p$ this becomes

$$\Omega^n I_n = (-1)^{n+1}(n-1)!\delta_s^{\ r}.$$

Thus if the operator Ω is applied n times to the n-rowed determinant $|E_s^{\ r}|$ we obtain $(-1)^{n+1}(n-1)!$ times the unit matrix.

TRINITY COLLEGE

* See *A note on the characteristic equation of a matrix*, American Mathematical Monthly, vol. 38 (1931), p. 386.

SOLUTIONS OF BOUNDED VARIATION OF THE FREDHOLM-STIELTJES INTEGRAL EQUATION*

BY L. J. PARADISO

The purpose of this note is to give conditions under which the Stieltjes integral equation

$$\phi(x) = f(x) + \lambda \int_a^b K(x, y)d\phi(y) \tag{1}$$

has a solution $\phi(x)$ of bounded variation.

In the first theorem conditions on $f(x)$ and $K(x, y)$ are given under which the method of successive substitutions yields a solution of bounded variation for a limited range of values of λ.

With further restrictions on $K(x, y)$, it is shown in the second theorem that the Fredholm method applies and the solution of bounded variation thus obtained is valid for all values of λ except for the characteristic values.

Finally, an example is given to show that the more restrictive conditions on $K(x, y)$ given in the second theorem are not sufficient to make the problem a special case of that treated by Riesz.†

THEOREM 1. *If*

(a) $f(x)$ *is of bounded variation,* $a \leqq x \leqq b$,

(b) $K(x, y)$, *defined and bounded on* $R(a \leqq x \leqq b,\ a \leqq y \leqq b)$, *is continuous in* y *for each* x *and has a total variation in* x *for each* y, $T_K(y)$, *which is a bounded function of* y *having the least upper bound* T_K *and*

$$|\lambda| < 1/T_K, \tag{c}$$

then the function $\bar{\phi}(x)$ *defined by the series*

$$\bar{\phi}(x) = f(x) + \lambda \int_a^b K(x, y_1)df(y_1) + \lambda^2 \int_a^b K(x, y_1)d \int_a^b K(y_1, y_2)df(y_2) + \cdots \tag{2}$$

is the unique solution of bounded variation of integral equation (1).

* Presented to the Society, December 31, 1930.

† F. Riesz, *Über lineare Funktionalgleichungen*, Acta Mathematica, vol. 41 (1918), pp. 71–98.

This theorem is proved by the usual method of successive substitutions. Let the function $\psi(x)$ be defined by

$$\psi(x) = \int_a^b K(x, y)df(y).$$

Designate the total variations of $f(x)$ and $\psi(x)$ by T_f and T_ψ respectively and let $M \geqq |K(x, y)|$. Then we have

$$|\psi(x)| \leqq MT_f \text{ and } T_\psi \leqq T_K \cdot T_f.$$

With the use of these inequalities it is found that series (2) obtained from equation (1) by repeated substitution converges absolutely and uniformly in x for all values of λ satisfying the inequality $|\lambda| < 1/T_K$. The function thus defined, $\bar{\phi}(x)$, is found by substitution to satisfy equation (1). That it is the unique solution follows as a consequence of the fact that the method of successive substitutions when applied to the homogeneous equation

$$\phi(x) = \lambda \int_a^b K(x, y)d\phi(y)$$

yields as its only solution $\phi(x) \equiv 0$.

THEOREM 2. *If $f(x)$ is of bounded variation in the interval (a, b) and $K(x, y)$ together with $\partial K(x, y)/\partial x$ are continuous functions of x and y in R, then there exists a unique solution $\phi(x)$ of bounded variation of equation* (1) *for all values of λ except for the characteristic values.*

This theorem can be proved by applying Fredholm's method to equation (1). However, we shall employ the following transformation* which reduces the problem to the solution of a Riemann integral equation. Let $\theta(x) = \phi(x) - f(x)$. Then equation (1) becomes

$$\theta(x) = \lambda \int_a^b K(x, y)df(y) + \lambda \int_a^b K(x, y)d\theta(y). \tag{3}$$

It is easily verified, with the given hypotheses on the functions involved, that each term of the right hand side of equation (3) is a continuous function of x and possesses a continuous

* J. D. Tamarkin suggested to me the possibility of transforming equation (1) into a Riemann integral equation.

derivative. Hence on placing $\lambda\int_a^b K(x, y)df(y) = F(y)$, we obtain from equation (3) by differentiation

$$\theta'(x) = F'(x) + \lambda \int_a^b \frac{\partial}{\partial x} K(x, y)\theta'(y)dy. \tag{4}$$

The function $F'(x)$ is continuous, and, moreover, by hypothesis $\partial K(x, y)/\partial x$ is continuous in x and y. Hence the Riemann integral equation (4) has a unique continuous solution $\theta'(x)$ for all values of λ except for the characteristic values. The unique solution of bounded variation $\phi(x)$ of equation (1) can thus be found.

An example. Let $\|f\|$ denote the maximum of the absolute value of the continuous function $f(x)$ in (a, b). One of the conditions on the transformation

$$T[f] = \int_a^b K(x, y)df(y)$$

as given by F. Riesz* is that there exists a constant M such that for all continuous functions $f(x)$

$$\|T[f]\| \leqq M\|f\|. \tag{5}$$

The function $K(x, y)$ defined in the following example satisfies the conditions of Theorem 2 whereas inequality (5) given by Riesz is not satisfied by the transformation $T[f]$. We define $K(x, y)$ to be a function of one variable, thus:

$$K(x, 0) = 0, \quad K(x, y) = y \sin(\pi/y), \ (0 < y \leqq 1).$$

We next define a sequence of continuous functions $f_n(y)$ bounded in n and y. The nth member of this sequence is a function whose graph consists of a series of broken lines. These lines have a slope equal to zero in the interval $(0, 1/(n+1))$, while in the interval $(1/(n+1), 1)$ they have a negative slope where the function $y \sin(\pi/y)$ is negative and a positive slope where this function is positive. Let δ_i denote 1 or 0 according as i is odd or even. Then

$$f_n(y) = (-1)^k k(k+1)y + (-1)^{k+1}k + \delta_k,$$
$$1/(k+1) \leqq y \leqq 1/k, \ (k = 1, 2, \cdots, n),$$

* F. Riesz, loc. cit., p. 72.

and

$$f_n(y) = \delta_n, \qquad (0 \leqq y \leqq 1/(n+1)).$$

From this definition we have

$$\frac{df_n(y)}{dy} = (-1)^k k(k+1), \qquad (1/(k+1) < y < 1/k,\ k = 1, 2, \cdots, n), \tag{6}$$

$$= 0, \qquad (0 < y < 1/(n+1)).$$

The transformation

$$T[f_n] = \int_0^1 y \sin(\pi/y) df_n(y), \quad (n = 1, 2, \cdots),$$

becomes from the definition of $f_n(y)$ and from (6)

$$T[f_n] = \sum_{k=1}^{k=n} \int_{1/(k+1)}^{1/k} y \sin(\pi/y)(-1)^k k(k+1)dy \tag{7}$$

$$= \sum_{k=1}^{k=n} k(k+1) \int_{1/(k+1)}^{1/k} (-1)^k y \sin(\pi/y)dy.$$

The $\int_{1/(k+1)}^{1/k} y \sin(\pi/y)\, dy$ is in absolute value greater than the area of the triangle whose vertices are at $(1/(k+1), 0)$, $(1/k, 0)$, $[2/(2k+1), (-1)^k 2/(2k+1)]$. Since the area of this triangle is $1/[k(k+1)(2k+1)]$, we have

$$\int_{1/(k+1)}^{1/k} (-1)^k y \sin(\pi/y)dy > 1/[k(k+1)(2k+1)].$$

Consequently we get from (7)

$$T[f_n] > \sum_{k=1}^{k=n} k(k+1)/[k(k+1)(2k+1)] = \sum_{k=1}^{k=n} 1/(2k+1),$$

from which it follows that $T[f_n]$ becomes infinite with n. Hence condition (5) is not satisfied and equation (1) is not a special case of that treated by F. Riesz.

CORNELL UNIVERSITY

ON REAL SYMMETRIC DETERMINANTS WHOSE PRINCIPAL DIAGONAL ELEMENTS ARE ZERO

BY W. V. PARKER

A paper by L. M. Blumenthal in this Bulletin* is concerned with the following theorem.

If the symmetric determinant

$$D = \begin{vmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & r_{12} & r_{13} & r_{14} \\ 1 & r_{21} & 0 & r_{23} & r_{24} \\ 1 & r_{31} & r_{32} & 0 & r_{34} \\ 1 & r_{41} & r_{42} & r_{43} & 0 \end{vmatrix}, \qquad (r_{ij} = r_{ji}),$$

with $r_{ij} > 0$, $(i, j = 1,2,3,4)$, $i \neq j$, *is different from zero, and the complementary minors of four of the elements in the principal diagonal vanish, then the complementary minor of the remaining element does not vanish.*

The purpose of this note is to establish a more general theorem from which the above is immediately obtainable. This theorem may be stated as follows.

THEOREM 1. *If the real symmetric determinant,* $D = |a_{ij}|$, *of order five with* $a_{ii} = 0$, $(i, j = 1,2,3,4,5)$, *has all five of its fourth-order principal minors zero, then* D *vanishes.*

Since the elements of D are real, we may write

$$D = \begin{vmatrix} 0 & a^2 & b^2 & c^2 & d^2 \\ a^2 & 0 & x^2 & y^2 & z^2 \\ b^2 & x^2 & 0 & u^2 & v^2 \\ c^2 & y^2 & u^2 & 0 & w^2 \\ d^2 & z^2 & v^2 & w^2 & 0 \end{vmatrix},$$

where a,b,c,d,x,y,z,u,v,w are either real or pure imaginary numbers. The theorem is trivial if all elements of the first row are zero. We will suppose then that at least one element is different from zero and without loss of generality we may assume $a \neq 0$.

* This Bulletin, vol. 37 (1931), pp. 752–758.

If now we subtract b^2/a^2 times the second column from the third, c^2/a^2 times the second column from the fourth, and d^2/a^2 times the second column from the fifth, and then transform the rows in the same way, D assumes the form

$$(1)\quad D=-\frac{1}{a^2}\begin{vmatrix} -2b^2x^2 & a^2u^2-b^2y^2-c^2x^2 & a^2v^2-d^2x^2-b^2z^2 \\ a^2u^2-b^2y^2-c^2x^2 & -2c^2y^2 & a^2w^2-d^2y^2-c^2z^2 \\ a^2v^2-d^2x^2-b^2z^2 & a^2w^2-d^2y^2-c^2z^2 & -2d^2z^2 \end{vmatrix}.$$

If we denote the fourth-order principal minors by M_1, M_2, M_3, M_4, M_5, we have

$$M_1=(xw+yv+zu)(xw+yv-zu)(xw-yv+zu)(xw-yv-zu),$$
$$M_2=(bw+du+cv)(bw+du-cv)(bw-du+cv)(bw-du-cv),$$
$$M_3=(aw+dy+cz)(aw+dy-cz)(aw-dy+cz)(aw-dy-cz),$$
$$M_4=(av+dx+bz)(av+dx-bz)(av-dx+bz)(av-dx-bz),$$
$$M_5=(au+cx+by)(au+cx-by)(au-cx+by)(au-cx-by).$$

If $M_3=M_4=M_5=0$, then $a^2u^2=(cx\pm by)^2$, $a^2v^2=(dx\pm bz)^2$, and $a^2w^2=(dy\pm cz)^2$. Substituting these values in (1) we get

$$(2)\quad D=-\frac{(8b^2c^2d^2x^2y^2z^2)}{a^2}\begin{vmatrix} -1 & \pm 1_u & \pm 1_v \\ \pm 1_u & -1 & \pm 1_w \\ \pm 1_v & \pm 1_w & -1 \end{vmatrix}$$
$$=2[(\pm 1_u)(\pm 1_v)(\pm 1_w)+1](-8b^2c^2d^2x^2y^2z^2)/a^2,$$

where $\pm 1_u$ is $+1$ or -1 according as $a^2u^2=(cx+by)^2$ or $a^2u^2=(cx-by)^2$ and similarly for $\pm 1_v$ and $\pm 1_w$. From this form it is evident that if b,c,d,x,y, or z is zero, D is zero and hence our theorem is true for this case. Let as assume now that b,c,d,x,y,z are all different from zero. Then in order that D be different from zero it is necessary that the above signs be taken *all plus* or *one plus and two minus*. Since M_1 and M_2 are even functions of u,v,w, we may take only positive square roots on both sides.

CASE I. *The three signs plus.* In this case $au=cx+by$, $av=dx+bz$, $aw=dy+cz$. Substituting these values in the expressions for M_1 and M_2 above we have

$$M_1=-(dxy+cxz+byz)(16bcdx^2y^2z^2)/a^4,$$
$$M_2=-(cdx+bdy+bcz)(16b^2c^2d^2xyz)/a^4.$$

Since by supposition b,c,d,x,y,z are all different from zero, in order that $M_1=M_2=0$ it is necessary that $dxy+cxz+byz=0$ and $cdx+bdy+bcz=0$, from which it follows that $c^2x^2+bcxy+b^2y^2=0$. That is $cx/(by)$ must be a complex cube root of unity, but this is impossible since $cx/(by)$ is either real or a pure imaginary number. Hence the theorem is true for this case.

CASE II. *One sign plus and two minus.* There are three possibilities here. We will carry the three through simultaneously indicating the corresponding steps by (1), (2), (3) for each case:

$$(1)\begin{cases} au = cx + by, \\ av = dx - bz, \\ aw = dy - cz, \end{cases} \quad (2)\begin{cases} au = cx - by, \\ av = dx + bz, \\ aw = dy - cz, \end{cases} \quad (3)\begin{cases} au = cx - by, \\ av = dx - bz, \\ aw = dy + cz. \end{cases}$$

The corresponding expressions for M_1 and M_2 are

$$(1) \qquad M_1 = -(dxy - cxz - byz)(16bcdx^2y^2z^2)/a^4,$$

$$(2) \qquad M_1 = -(cxz - dxy - byz)(16bcdx^2y^2z^2)/a^4,$$

$$(3) \qquad M_1 = -(byz - dxy - cxz)(16bcdx^2y^2z^2)/a^4,$$

$$(1) \qquad M_2 = -(bcz - bdy - cdx)(16b^2c^2d^2xyz)/a^4,$$

$$(2) \qquad M_2 = -(bdy - bcz - cdx)(16b^2c^2d^2xyz)/a^4,$$

$$(3) \qquad M_2 = -(cdx - bcz - bdy)(16b^2c^2d^2xyz)/a^4.$$

Since b,c,d,x,y,z are all different from zero, in order that $M_1=M_2=0$ it is necessary that

$$(1) \qquad c^2x^2 + bcxy + b^2y^2 = 0,$$

$$(2) \qquad d^2x^2 + bdxz + b^2z^2 = 0,$$

$$(3) \qquad d^2y^2 + cdyz + c^2z^2 = 0.$$

None of these is possible for the same reason as in Case I.

We have shown therefore that if $M_3=M_4=M_5=0$ and $D\neq 0$, then not both M_1 and M_2 are zero. Hence, if $M_1=M_2=M_3=M_4=M_5=0$, D vanishes.

If the elements outside the principal diagonal of D are all different from zero, we see from (2) that if $M_3=M_4=M_5=0$ then the signs must be taken *all minus* or *two plus and one minus* in order for D to vanish. In each of these cases $M_1=M_2=0$. We may state, in this case, therefore the following corollary.

COROLLARY. *If four of the principal minors of D are zero and the fifth is not zero then D is not zero.*

As an immediate consequence of Theorem 1 we have also the following result.

THEOREM 2. *If $D=|a_{ij}|$ is a real symmetric determinant of order greater than four with $a_{ii}=0$, then if all fourth-order principal minors of D are zero, D vanishes.*

Any fifth-order principal minor of D is of the form of the determinant under Theorem 1 and hence is zero if all fourth-order principal minors vanish. Hence all principal minors of orders four and five are zero and therefore the rank of D is three or less.*

THEOREM 3. *If $D=|a_{ij}|$ is a real symmetric determinant of order n, $n>5$, with $a_{ii}=0$, and M is any principal minor of D of order $n-1$, then if all fourth-order principal minors of M are zero, D vanishes.*

By Theorem 2, the rank of M is three or less. We may suppose now that M is the minor in the upper left hand corner of D. When D is expanded according to the product of the elements of its last row and last column it is expressed as a sum of products of these elements by minors of M of order $n-2$†. But all minors of M of order $n-2$ are zero and hence D vanishes.

MISSISSIPPI WOMAN'S COLLEGE

* Bôcher, *Introduction to Higher Algebra*, p. 57, Theorem 2.

† Bôcher, loc. cit., p. 29, Theorem 3.

ON THE DEGREE OF GENERALITY OF A CLASS OF ARITHMETICAL IDENTITIES

BY E. T. BELL

1. *Introduction.* In a recent paper, *A class of arithmetical identities*,* Oppenheim, in commenting on a special case (n not square) of the identity A below, which is due to Uspensky†, states that the conditions on the function H are "needlessly restrictive," and he gives a "general identity which will include such identities as these," (A, with n not square, and the new identities of Uspensky which involve incomplete numerical functions).‡ It is indeed readily seen that *Oppenheim's identity* B *implies all of those of Uspensky.* In view of the remarks quoted, it is perhaps in order to give a detailed proof, by general methods adapted to the specific functions and partitions in A, B, that *Uspensky's* A *alone implies Oppenheim's* B. *The identities* A, B *are thus of precisely the same degree of generality.*

2. *The Identities.* Each of the functions $H(w, u, v)$, $F(w, u, v)$, $\phi(w, u, v)$ is finite and single-valued for integer values of w, u, v, and *beyond this restriction and those stated in* (1), (4), (5) *for* H, F ϕ, *respectively, the functions are entirely arbitrary.*

A. (USPENSKY). *Let*

$$(1)\quad H(w, u, v) = -H(-w, u, v) = H(w, -u, -v); H(0, u, v) = 0.$$

Then, if the $\sum$ *on the left of* (3) *refers to all integers* x, y, z *such that*

$$(2)\qquad x^2 + yz = n, \quad y > 0, \quad z > 0,$$

where n *is an arbitrary constant integer,*

---

* Quarterly Journal of Mathematics, Oxford Series, vol. 2, No. 7, 1931, pp. 300–303.

† First published, apparently, in Kharkof Mathematical Society Communications, (2), vol. 15 (1915–1917), pp. 81–147, (Russian). See Jahrbuch über die Fortschritte der Mathematik, vol. 48, p. 1350.

‡ This Bulletin, vol. 36 (1930), pp. 743–754. Oppenheim, loc. cit., p. 230, corrects the misprints. The considerations of the present note apply also to the identities of Mordell, which Oppenheim cites.

$$(3)\quad \sum[H(y+z, z-y, z) - 2H(z-2x, 2x+2y-z, x+y)] = \epsilon(n)\sum_{j=1}^{2m-1}[H(2m, 2m-2j, m-j) - 2H(2m-j, 2m-j, m)],$$

where $\epsilon(n)=1$ *or* 0 *according as* n *is or is not square, and* $m \equiv \lfloor n^{1/2} \rfloor$.

B. (OPPENHEIM). *Let*

$$(4)\quad F(w, u, v) + F(w, -u, -v) + F(-w, u, v) + F(-w, -u, -v) = 0,$$

$$(5)\quad \phi(w, u, v) = \phi(-w, u, v) = \phi(w, -u, -v).$$

Then

$$(6)\quad \sum F(y+z, z-y, x) = \sum[F(\xi, \eta, \zeta) + F(\xi, -\eta, -\zeta)],$$

where $\sum$ *on the left refers to all integers* x, y, z *such that*

$$(7)\quad \phi(y+z, z-y, x) = N,\ y>0,\ z>0,\ \mu(z-y) \neq \nu x,$$

where N *is an arbitrary (real or complex) constant number,* μ, ν *are any real constant numbers, and* $\sum$ *on the right refers to all integers* ξ, η, ζ *such that*

$$(8)\quad \phi(\xi, \eta, \zeta) = N,\ 0 \neq \xi + \eta \equiv 0 \bmod 2,\ \xi > \eta,\ \mu\eta > \nu\zeta.$$

If $\phi = N$ *has an infinity of solutions,* F *is assumed to be such that the infinite series concerned are absolutely convergent.*

In a previous note* I showed that A implies and is implied by the identity

$$\begin{vmatrix} a_3 & b_3 & c_3 \\ a_2 & b_2 & c_3 \\ a_2 & b_2 & c_2 \end{vmatrix} = 0,$$

$$a_j = \theta_j(2x+y+2x),\quad b_j = \theta_j(2x-y-2z),\quad c_j = \theta_j(y).$$

The connection of this and the present note with the series† on *Quadratic partitions* in course of publication in this Bulletin is indicated below, in §4.

3. *Implication of* B *by* A. By the definition of H, we may replace H in (3) by H_1, where $H_1(w, u, v)$ is *any* function which

* This Bulletin, vol. 32 (1926), pp. 682–688, especially p. 685.

† See, for example, this Bulletin, vol. 37, pp. 870–875.

is single-valued and finite for integer values of w, u, v, and which satisfies the parity conditions (1). We shall take

$$H_1(w, u, v) \equiv \Phi_N(w, u, v)\Psi_{\mu,\nu}(u, v)H(w, u, v), \tag{9}$$

where, $\Phi_N(w, u, v) = 1$ or else $= 0$, according as $\phi(w, u, v) = N$ or $\phi(w, u, v) \neq N$ (N as in B), and where $\Psi_{\mu,\nu}(u, v) = 1$ or 0 according as $\mu u - \nu v \neq 0$ or $\mu u - \nu v = 0$. This choice is permissible, since

$$\Phi_N(w, u, v) = \Phi_N(-w, u, v) = \Phi_N(w, -u, -v),$$
$$\Psi_{\mu,\nu}(u, v) = \Psi_{\mu,\nu}(-u, -v),$$

as is evident from the definition of the functions. Having made the above substitution in A, we then replace $H(w, u, v)$ in the resulting identity (or at once in (9), before making the substitution), *by its instance* $F(w, u, v) + F(w, -u, -v)$, where F is as in (4). Thus A implies

$$\begin{aligned}(10)\qquad &\sum_1 [F(y+z, z-y, x) + F(y+z, -z+y, -x)] \\ &- 2\sum_2 [F(z-2x, 2x+2y-z, x+y) \\ &+ F(z-2x, -2x-2y+z, -x-y)] \\ &= \epsilon(n)\sum_3 [F(2m, 2m-2j, m-j) + F(2m, -2m+2j, -m+j)] \\ &- 2\epsilon(n)\sum_4 [F(2m-j, 2m-j, m) \\ &+ F(2m-j, -2m+j, -m)],\end{aligned}$$

where the limits $j = 1$ to $j = 2m-1$ are to be supplied to $\sum_3, \sum_4$, and $\sum_i$ refers in an obvious way to all the *simultaneous* integer solutions of (2) and (11.i),

$$(11.1)\qquad \phi(y+z, z-y, x) = N, \qquad \mu(z-y) \neq \nu x;$$
$$(11.2)\qquad \phi(z-2x, 2x+2y-z, x+y) = N, \quad \mu(2x+2y-z) \neq \nu(x+y);$$
$$(11.3)\qquad \phi(2m, 2m-2j, m-j) = N, \quad \mu(2m-2j) \neq \nu(m-j);$$
$$(11.4)\qquad \phi(2m-j, 2m-j, m) = N, \qquad \mu(2m-j) \neq \nu m.$$

By (2), (5), (11.1) the $\sum_1$ term in (10) is equal to

$$2\sum_1 F(y+z, z-y, x). \tag{12}$$

By (5), if $(w, u, v) = (W, U, V)$ is an integer solution of $\phi(w, u, v) = N$, then so also is (aW, bU, bV), where each of a, b is either of $1, -1$, and for *any definite* (a, b) as (w, u, v) runs over all solutions, so also does (aW, bU, bV). Apply this to (11.3), (11.4) with $(a, b) = (-1, 1)$. Then $\sum_i$ in (10), for $i = 3, 4$, is half what $\sum_i$ becomes when $F(w, u, v)$ is replaced by $F(w, u, v) + F(-w, u, v)$. Hence, by (4),

$$\sum_3 = 0, \quad \sum_4 = 0. \tag{13}$$

Consider now the set S of all solutions (x, y, z) of (7), and let S_n be that subset of S which contains all those solutions, and only those, which satisfy (2). Then S_p, S_q are mutually exclusive if $p \neq q$, and S is the logical sum of $S_n (n = 1, 2, \cdots)$. Any particular S_n may be the null set. In (10) take $n = 1, 2, \cdots$, and add corresponding members of all the resulting identities. Then, since (10) is subject to (2), (11.1)-(11.4), it follows by the above and (12), (13) that A implies

$$\begin{aligned} &\sum F(y + z, z - y, x) \\ &= \sum [F(z - 2x, 2x + 2y - z, x + y) \\ &\quad + F(z - 2x, -2x - 2y + z, -x - y)], \end{aligned} \tag{14}$$

where $\sum$ in the first line refers to (7), and $\sum$ in the second line to (11.2) with the additional conditions $y > 0, z > 0$.

To reduce (14) to the form (B), we make the substitution

$$(z - 2x, 2x + 2y - z, x + y) = (a\xi, b\eta, b\zeta), \tag{15}$$

where each of a, b is a definite one of $1, -1$. Then

$$(2x, 2y, z) = (2b\zeta - a\xi - b\eta, a\xi + b\eta, 2b\zeta - b\eta). \tag{16}$$

By the derivation of (14), it is sufficient to consider (14) subject to (2). Since y, z are integers and $y > 0, z > 0$, we see by (16) that the right member of (14) in the case considered is equal to

$$\sum [F(\xi, \eta, \zeta) + F(\xi, -\eta, -\zeta)], \tag{17}$$

the $\sum$ referring to all integer solutions ξ, η, ζ of $\phi(\xi, \eta, \zeta) = N$ which are such that

$$n = (b\zeta)^2 + \tfrac{1}{4}(a\xi + b\eta)(a\xi - b\eta), \tag{18}$$

$$a\xi + b\eta \equiv 0 \mod 2, \tag{18.1}$$

$$a\xi + b(\eta) > 0, \tag{18.2}$$

$$b(2\zeta - \eta) > 0, \tag{18.3}$$

$$b(\mu\eta - \nu\zeta) \neq 0, \tag{18.4}$$

for any definite (a, b) as defined.

Let us now replace (18.4) by its equivalent

$$b(\mu\eta - \nu\zeta) > 0 \text{ or } b(\mu\eta - \nu\zeta) < 0. \tag{19}$$

In the first of (19) take $b=1$; in the second $b=-1$. Then the sum (17) in the case being considered is *half* the same sum over all integers ξ, η, ζ such that $\phi(\xi, \eta, \zeta) = N$, subject to the sets of conditions (in the two columns)

$$n = (+\zeta)^2 + \tfrac{1}{4}(a\xi + \eta)(a\xi - \eta), \qquad n = (-\zeta)^2 + \tfrac{1}{4}(a\xi - \eta)(a\xi + \eta), \tag{20}$$

$$a\xi + \eta \equiv 0 \bmod 2, \qquad a\xi - \eta \equiv 0 \bmod 2, \tag{20.1}$$

$$a\xi + \eta > 0, \qquad a\xi - \eta > 0, \tag{20.2}$$

$$2\zeta - \eta > 0, \qquad 2\zeta - \eta < 0, \tag{20.3}$$

$$\mu\eta - \nu\zeta > 0, \qquad \mu\eta - \nu\zeta > 0. \tag{20.4}$$

The conditions (20.3) may be suppressed, since the one missing possibility, $2\zeta - \eta = 0$, contributes zero to (17). This is seen as follows. If $2\zeta = \eta$, then by (20.1), ξ is even; by (20), n is square, $a\xi = \pm 2m$ (m as in A). Choose $a=1$, or $a=-1$, according as $\xi > 0$, or $\xi < 0$. Adding half the respective contributions to (17), we see by (4) that the total contribution vanishes.

Combining all results so far, taking $a=1$ in both columns of (20)–(20.4), and summing over both for all values of n, we see that A implies B with the exception that the condition $\xi + \eta \neq 0$ in (8) is here replaced by $\xi + \eta > 0$. To complete the proof that A implies B it is therefore sufficient to show that if $\xi + \eta \neq 0$ be replaced by $\xi + \eta < 0$ in (8), then the right of (6) vanishes. This follows by (4), since the conditions $\xi > \eta$, $\xi + \eta < 0$, $\mu\eta > \nu\zeta$ are invariant under the substitution $(\xi', \eta', \zeta') = (-\xi, \eta, \zeta)$, where ξ, η, ζ are subject to the conditions (8) with $\xi + \eta \neq 0$ replaced by $\xi + \eta < 0$, and since (ξ', η', ζ') runs over all solutions of $\phi(\xi, \eta, \zeta) = N$ when (ξ, η, ζ) does, by (5).

4. *Parity Transformations.* The equivalence of A, B is a simple instance of the invariance of a parity identity under a parity transformation.* Such a transformation leaves invariant the parity of the functions in the identity, or that of the given partition, or both. In this instance both are used; in (9) and in passing to (10) the transformation is on the function; in substituting

* For some account of these, see E.T. Bell, *Algebraic Arithmetic*, Colloquium Publications of this Society, vol. 7, 1927, and earlier papers in the Transactions of this Society, where such transformations were applied in simple cases to the arithmetic of higher forms.

$\phi(w, u, v)$ for w^2+uv, in which $\phi(w, u, v)$ and w^2+uv are of the same parity, it is on the partition.

If we define (as seems reasonable) the proposition P to be a generalization of the proposition Q when and only when P implies Q, and Q does not imply P, it can be proved that it is impossible to generalize a parity identity by parity transformations.* For this reason, in the notes appearing currently in this Bulletin on *Quadratic partitions*, the discussion is limited to such partitions and to parity functions which, so far as feasible, are unrestricted by conditions which can be introduced by parity transformations. The passage from n to N also is a parity transformation, since any not identically zero linear homogeneous function of functions having a given parity has the same parity.

The linear transformation (15) would be suggested naturally as a possible means of transforming A by that which is given by the addition theorem for the thetas used in deducing A from the identity stated in §2, end. All such transformations go back ultimately to the orthogonal one on four variables which makes possible Jacobi's theta formula. It can be shown (and will be discussed in another note) from theorems of Dickson† on certain automorphs of sums of like powers, higher than the second, that parity identities referring to partitions of degree higher than 2 not obtainable by parity transformations from quadratic partitions do not exist.

Finally it may be noticed that convergence is irrelevant if parity identities be given their most elementary significance, that of equalities between *sets* of vectors having parity. This was observed essentially by Glaisher,‡ although he did not state it explicitly.

California Institute of Technology

* The like applies to the broader class of linear homogeneous transformations of the *coordinates* in the *vector* (or *matric*) variables of a parity function, although the parity of the *transformed* function, considered as a function of the *new* variables, is in general different from that of the function in the *old* variables. All such transformations amount merely to linear homogeneous transformations of the indeterminates in the trigonometric identity which is equivalent to the parity identity.

† Mathematische Annalen, vol. 52 (1899), pp. 561–582.

‡ Proceedings of the London Mathematical Society, vol. 22 (1890–91), pp. 359–410. In Dickson's *History*, this work is reported in vol. 1, p. 310, under *Sum and Number of Divisors*. It might have been cited also in vol. 2, Chap. 11.

ON THE MOMENT PROBLEM FOR A FINITE INTERVAL

BY T. H. HILDEBRANDT

Hausdorff* and Schoenberg† have shown that a necessary and sufficient condition that there exist a function of bounded variation $\chi(t)$, such that we have, for every n, $\mu_n = \int_0^1 t^n d\chi(t)$, is that $\sum_{m=0}^{n} \binom{n}{m} |\Delta^{n-m}\mu_m|$ form a bounded sequence. The fact that if such a function $\chi(t)$ exists, then $\Delta^m\mu_n = \int_0^1 t^n(1-t)^m d\chi(t)$, suggests that perhaps the sufficiency part of this theorem can be easily demonstrated by using the Bernstein polynomials and the Riesz theorem on linear functional operations on the set of continuous functions.

For any function $f(t)$ on (0, 1) the nth Bernstein polynomial is defined by the formula

$$B_n(f) = \sum_{m=0}^{n} \binom{n}{m} f(m/n) t^m (1-t)^{n-m},$$

and it is well known that the polynomials $B_n(f)$ converge uniformly to $f(t)$ if f is continuous. Further, it can be easily shown by induction‡ that if $P_k(t)$ is any polynomial of degree k, then

$$B_n(P_k) = P_k(t) + \sum_{i=1}^{k-1} A_{ik}(t)/n^i,$$

where $A_{ik}(t)$ are polynomials of degree at most k and are independent of n.

Define now a linear operation U on polynomials $P_k(t)$ by the condition

$$U(P_k) = U(a_0 + a_1 t + \cdots + a_k t^k) = \sum_{i=0}^{k} a_i \mu_i.$$

---

* Mathematische Zeitschrift, vol. 16 (1923), p. 238.

† This Bulletin, vol. 38 (1932), p. 72.

‡ Following the suggestion in Pólya and Szegö, *Aufgaben und Lehrsätze aus der Analysis*, vol. I, p. 6, Ex. 40; and p. 66, Ex. 144.

Then for any function f,

$$U(B_n(f)) = \sum_{m=0}^{n} f(m/n)\binom{n}{m}\Delta^{n-m}\mu_m.$$

As a consequence, for every n,

$$|\,U(B_n(f))\,| \leqq F \cdot M,$$

where F is the least upper bound of $|f(t)|$ on (0, 1) and M is the least upper bound of $\sum_{m=0}^{n}\binom{n}{m}|\Delta^{n-m}\mu_m|$. Hence if P_k is any polynomial of degree k, then

$$|\,U(P_k)\,| \leqq |\,U(B_n(P_k))\,| + \sum_{i=0}^{k-1} |\,U(A_{ik})\,|/n^i,$$

so that

$$|U(P_k)| \leqq \max|P_k| \cdot M.$$

It follows from this inequality that if f is any continuous function on (0, 1), then the sequence of numbers $U(B_n(f))$ satisfy the Cauchy condition of convergence, that is, converge to a finite limit, which limit we denote by $U(f)$. Obviously $U(f)$ is linear in f and $|U(f)| \leqq F \cdot M$. Hence by the Riesz theorem on linear limited operations, there exists a function $\chi(t)$ of bounded variation, such that

$$U(f) = \int_0^1 f(t)d\chi(t), \quad \text{that is,} \quad \mu_n = \int_0^1 t^n d\chi(t).$$

THE UNIVERSITY OF MICHIGAN

NEW DIOPHANTINE AUTOMORPHISMS*

BY LEONARD CARLITZ†

1. *Introduction.* In this note we consider the construction of an infinite set of homogeneous polynomials f such that

$$f(Y_1, \cdots, Y_n) = f^h(y_1, \cdots, y_n), \tag{1}$$

where

$$Y_i = Y_i(y_1, \cdots, y_n), \qquad (i = 1, \cdots, n), \tag{2}$$

are homogeneous of degree h. In that case, f is said to admit of a diophantine automorphism.‡ The construction depends on a very simple principle connected with invariant theory.

Let ϕ be a binary form of degree δ:

$$\phi = a_x^\delta = a_0x_1^\delta + \binom{\delta}{1}a_1x_1^{\delta-1}x_2 + \cdots + \binom{\delta}{\delta}a_\delta x_2^\delta;$$

let q be a quadratic covariant of ϕ of degree ρ; then ψ, the discriminant of q, is an invariant of the ground-form ϕ of degree 2ρ. By the property of invariance, if under the linear transformation

$$x_1 = \alpha_{11}X_1 + \alpha_{12}X_2, \quad x_2 = \alpha_{21}X_1 + \alpha_{22}X_2, \tag{3}$$

of determinant $\Delta = \alpha_{11}\alpha_{22} - \alpha_{21}\alpha_{12}$, the form ϕ becomes

$$\Phi = A_X^\delta = A_0X_1^\delta + \binom{\delta}{1}A_1X_1^{\delta-1}X_2 + \cdots,$$

and ψ becomes $\Psi = \psi(A)$, then

$$\psi(A) = \Delta^{\delta\rho}\psi(a). \tag{4}$$

Now putting $q = q_{11}x_1^2 + 2q_{12}x_1x_2 + q_{22}x_2^2$, we take in (3)

$$\alpha_{11} = q_{12}, \quad \alpha_{12} = q_{22}, \quad \alpha_{21} = -q_{11}, \quad \alpha_{22} = -q_{12}.$$

Evidently (4) becomes

---

* Presented to the Society, April 11, 1931.

† National Research Fellow.

‡ E. T. Bell, this Bulletin, vol. 33 (1927), pp. 71–80.

$$\psi(A) = \psi^{\delta\rho+1}(a), \tag{5}$$

where, by virtue of the specialized transformation (3),

$$A_i = A_i(a_0, \cdots, a_\delta) \tag{6}$$

are homogeneous of degree $\delta\rho+1$.

Clearly (5) and (6) furnish a solution of (1) and (2). And (for $\psi \neq$ square) we can assert immediately that it is surely non-trivial if δ is odd. Indeed the solution is trivial only if (6) reduces to

$$A_i = \alpha\psi^{\delta/2}(a) \cdot a_i.$$

2. *Expression of q Symbolically.* It is here convenient to make use of the symbolism of the theory of invariants.* Let us set

$$\phi = a_x^\delta = (a_1x_1 + a_2x_2)^\delta, \quad q = q_x^2 = (q_1x_1 + q_2x_2)^\delta,$$

so that upon applying (3), in specialized form, we obtain

$$\begin{aligned} a_x \to A_X &= a_1(q_1q_2X_1 + q_2^2X_2) + a_2(-q_1^2X_1 - q_1q_2X_2) \\ &= (a_1q_2 - a_2q_1)(q_1X_1 + q_2X_2) = (aq)q_X, \end{aligned} \tag{7}$$

and therefore ϕ becomes

$$\Phi = (aq)(aq') \cdots (aq^{(\delta-1)})q_Xq_X' \cdots q_X^{(\delta-1)},$$

where $q, q', \cdots$ are equivalent symbols. Similarly

$$q_x \to q_1(q_1'q_2'X_1 + q_2'^2X_2) + q_2(-q_1'^2X_1 - q_1'q_2'X_2) = (qq')q_X',$$

and therefore q becomes

$$Q = (qq')(qq'')q_X'q_X'' = \psi q, \tag{8}$$

as follows immediately from the fundamental identity of invariant theory.

3. *Consequences of* (7) *and* (8). The equations (7) and (8) imply an interesting result concerning the transformation (6). Clearly all the operations performed on ϕ may be performed on Φ. By (8), the transformation (3), properly specialized, changes (7) to

$$\psi^{\delta\rho/2} \cdot (aq)(qq')q_x'^{\cdots} = \tfrac{1}{2}\psi^{\delta\rho/2}(qq')^2a_x = \psi^{(\delta\rho+2)/2}a_x,$$

* See Grace and Young, *Algebra of Invariants*, 1903, Chap. 1.

and

$$\Phi \rightarrow \psi^{\delta(\delta\rho+2)/2}\phi;$$

that is, Φ is transformed into ϕ multiplied by a factor free of x. This is equivalent to

$$A_i(A) = \psi^{\delta(\delta\rho+2)/2} \cdot a_i, \tag{9}$$

where the exponent is surely integral, for when δ is odd ρ must be even. It follows that the equation (6) *defines a Cremona transformation of period two.*

Now from (9)

$$\left|\frac{\partial A_i(A)}{\partial A_j}\right| \cdot \left|\frac{\partial A_i}{\partial a_j}\right| = \left|\frac{\partial(\psi^{\delta(\delta\rho+2)/2}a_i)}{\partial a_j}\right|,$$

since the determinants involved are jacobians. But the jacobian on the right is easily shown to be equal to

$$(\delta\rho + 1)^2\psi^{\delta(\delta+1)(\delta\rho+2)/2}.$$

Hence, if ψ is algebraically irreducible,

$$\left|\frac{\partial A_i}{\partial a_j}\right| = c\psi^{\delta(\delta+1)/2},$$

where c denotes a numerical constant.

4. *Special Values of* δ. We now consider some special values of δ. If δ be odd, it is evident, to begin with, that a q is always furnished by the covariant

$$K_{\delta-1} = (\phi, \phi)^{\delta-1},$$

the $(\delta-1)$th transvectant of ϕ with itself. The quantity $K_{\delta-1}$ is an irreducible covariant, and its discriminant

$$\psi = \tfrac{1}{2}(K_{\delta-1}, K_{\delta-1})^2$$

is an irreducible invariant of degree four. By what precedes, the form ψ admits of a non-trivial automorphism.

Take $\delta = 3$. Then K_2 is the Hessian of ϕ, and ψ is the discriminant of ϕ as well as of K_2. By (7) we must here consider the transformed form

$$(aq)(aq')(aq'')q_x q_x' q_x''$$
$$= \tfrac{1}{2}(aq)q_x((aq')^2 q_x''^2 + (aq'')^2 q_x'^2 - (q'q'')^2 a_x^2)$$
$$= q\cdot(aq)(aq')^2 q_x - \psi\cdot(aq)a_x^2 q_x.$$

But the first term involves $((\phi, q)^2, q)^1$ and $(\phi, q)^2$ vanishes, as can be seen directly or by making use of the known result that the cubic has no linear covariants. Hence the transformed form is, but for a numerical factor, the product of the discriminant into the cubicovariant. In this case, then, (6) reduces to

$$\frac{1}{\psi} A_i = f_i(a_0, \cdots, a_3),$$

where f_i is of degree three.

5. *The Case* $\delta = 5$. When $\delta = 5$, there are three quadratic covariants in the irreducible set.* The simplest is the form $K_4 = (\phi, \phi)^4$, which is generally denoted by i. For the transformed form (7) we have here $(i = i_x^2 = i_x'^2 = \cdots)$

$$(10)\quad (ai)(ai')\cdots(ai^{iv})i_x i_x'\cdots i_x^{iv}$$
$$= \tfrac{1}{4}(ai)i_x(2(ai')^2\cdot i - Aa_x^2)(2(ai'')^2\cdot i - Aa_x^2)$$
$$= (ai)(ai')^2(ai'')^2 i_x\cdot i^2 + AF,$$

where $A = (i, i)^2$ is used in place of ψ, and F is an (integral) covariant whose exact form is immaterial. Now

$$(ai)(ai')^2(ai'')^2 i_x = (((\phi, i)^2, i)^2, i)^1$$
$$= (-(j, i)^2, i)^1 = -(\alpha, i)^1 = -\beta,$$

where j, α and β are irreducible covariants of order 3, 1, 1, respectively.† It is then clear that (10) is not a multiple of A and therefore in this case the transformation (6) is not reducible; that is, it is actually a Cremona transformation of degree $2\cdot 5+1 = 11$.

We have defined j above as $-(\phi, i)^2$; the second transvectant of j with itself is a second irreducible quadri-covariant, $\tau = (j, j)^2$. We now have, in place of (10),

---

* See Grace and Young, loc. cit., p. 131.

† See Grace and Young, loc. cit.

$$(11)\quad (a\tau)\cdots(a\tau^{\text{iv}})\tau_x\cdots\tau_x^{\text{iv}}$$
$$= \tfrac{1}{4}(a\tau)\tau_x(2(a\tau')^2\tau - Ca_x^2)(2(a\tau'')^2\tau - Ca_x^2)$$
$$= (a\tau)(a\tau')^2(a\tau'')^2\tau_x + CF,$$

where $C=(\tau,\tau)^2$ replaces ψ. But

$$(a\tau)(a\tau')^2(a\tau'')^2\tau_x = (((\phi,\tau)^2,\tau)^2,\tau^1),$$
$$(\phi,\tau)^2 = -i\alpha - \tfrac{2}{3}A_j.$$

It is easy to show, using the tables of transvectants in Grace and Young, that $(i\alpha,\tau)^2$ is expressible linearly in terms of $B\alpha$ and δ (B is an invariant, α and δ are linear covariants in the irreducible set of concomitants of ϕ); $(B\alpha,\tau)^1$ and (δ,τ) are expressible in terms of $C\beta$ and $B\gamma$, γ a third linear covariant. It is then evident that (11) is not divisible by C, and therefore the transformation (6) is here also of maximum degree $(6\cdot 5+1=31)$.

The remaining quadratic covariant $\theta=(i,\tau)$ may be treated in exactly the same way, and again it appears that (6) does not reduce to a transformation of lower degree.

6. *The Case* $\delta=7$. When $\delta=7$, there are several quadratic covariants. We limit ourselves to the $K_6=(\phi,\phi)^6=q$, $\psi=(q,q)^2$. For this case, (7) becomes

$$(12)\ (aq)\cdots(aq^{\text{vi}})q_x\cdots q_x^{\text{vi}} = (aq)(aq')^2(aq'')^2(aq''')^2q_x\cdot q^3 + \psi F.$$

Now $(\alpha q)(\alpha q')^2(\alpha q'')^2(\alpha q''')^2q_x=(\phi,q^4)^7$ and this is a member of the irreducible set of concomitants of the septimic.* Accordingly (12) is not divisible by ψ, and again (6) is of maximum degree $(2\cdot 7+1=15)$.

In general, if δ be odd $(=2k+1)$ and greater than three, then for $q=K_{\delta-1}$, (7) reduces to

$$(\phi,q^{k+1})^{2k+1}\cdot q^{k-1} + \psi F,$$

and it seems likely that this is not divisible by ψ, so that (6) would in this case always be of maximum degree $=2\delta+1$.

7. *Even Values of* δ. When δ is even it is necessary to consider covariants of somewhat higher degree. Thus for $\delta=6$ the simplest q is†

---

* See F. von Gall, Mathematische Annalen, vol. 31 (1888), p. 320.

† See Grace and Young, loc. cit., p. 156.

$$q = l = (\phi, i)^4, \quad i = (\phi, \phi)^4.$$

Then for (7) we get as above

$$(al)^2(al')^2(al'')^2 \cdot l^3 + \psi F,$$

where $\psi = (l, l)^2$. But

$$(al)^2(al')^2(al'')^2 = (\phi, l^3)^6,$$

which is a member of the irreducible set of concomitants distinct from ψ.

8. *Conclusion.* The method outlined in §1 is by no means restricted to binary forms ψ. It is indeed obvious that starting with any quadratic covariant of a form ϕ in any number of variables we may arrive at equations (5) and (6). The transformation involved will certainly not be trivial if the degree of ϕ is not divisible by the number of variables.*

An obvious instance of this is furnished by quadratic forms, and indeed it is sufficiently obvious that their discriminants have the automorphic property. Another simple though less obvious instance is furnished by the quadri-covariant of the quaternary cubic ($\phi = a_x^3 = b_x^3 = \cdots$):

$$(abcd)(abce)(adef)(bdef)c_x f_x.$$

By the remark made in the last paragraph, this will certainly not lead to a trivial transformation.

CALIFORNIA INSTITUTE OF TECHNOLOGY

* It is of course assumed that the discriminant of the quadri-covariant of ϕ is not a kth power, k being the number of variables.

NOTE ON THE DIOPHANTINE EQUATION

$$ax^2 + by^2 + cz^2 + dt^2 = 0$$

BY L. J. MORDELL

Suppose that the constants a, b, c, d are integers none of which is zero, and that

I. *a, b, c, d have no squared factors*;

II. *no three of a, b, c, d have a common factor, that is,* $[a, b, c] = 1$, *etc.*

Then, if we suppose the restrictions I, II apply throughout the paper, the equation

$$ax^2 + by^2 + cz^2 + dt^2 = 0 \tag{1}$$

has integer solutions not all zero, if, and only if,

III. *a, b, c, d are not all of the same sign,*

IV. *every odd prime factor p_{ab} of $[a, b]$ for which*

$$\left(\frac{-cd}{p_{ab}}\right) = -1, \tag{2}$$

must also satisfy

$$\left(\frac{-ab/p_{ab}^2}{p_{ab}}\right) = 1, \tag{3}$$

together with five corresponding conditions derived by permuting the letters. The symbol is that of quadratic residuacity.

V. (1) *Either* $abcd \equiv 2, 3, 5, 6, 7 \pmod 8$,

V. (2) *or* $abcd \equiv 1 \pmod 8$, $a+b+c+d \equiv 0 \pmod 8$;

these mean that we must exclude

$$a \equiv b \pmod 8, \qquad c \equiv d \pmod 8, \qquad a \equiv c \pmod 4,$$

and the corresponding sets derived by permuting the letters;

V. (3) *or two of a, b, c, d are even, say a, b, and we must have*

V. (3.1) *either* $\frac{1}{4}abcd \equiv 3, 5, 7 \pmod 8$,

V. (3.2) *or* $\frac{1}{4}abcd \equiv 1 \pmod 8$, and $\frac{1}{2}a+\frac{1}{2}b+c+d \equiv (c^2d^2-1)/2 \pmod 8$; these mean that we must exclude the sets

$$\frac{1}{2}a \equiv \frac{3}{2}b \pmod 8, \quad c \equiv 3d \pmod 8,$$

or

$$\frac{1}{2}a \equiv \frac{1}{2}b \pmod 8, \quad c \equiv d \pmod 8, \quad \frac{1}{2}a \equiv c \pmod 8,$$

or

$$\frac{1}{2}a \equiv \frac{5}{2}b \pmod 8, \quad c \equiv 5d \pmod 8, \quad \frac{1}{2}a \equiv -c \pmod 8.$$

If in equation (1), we impose the restrictions that

$$[z, t, a, b] = 1 \text{ or } 2 \tag{4}$$

together with the five similar ones obtained by permuting the letters, the only difference in the conditions is that IV must be replaced by

VI. *$-cd$ is a quadratic residue of $[a, b]$, written as $-cd\ R\ [a, b]$, and so of every odd prime factor of $[a, b]$; together with the five conditions obtained by permuting the letters.*

In a recent number of this Bulletin, R. G. Archibald* deduces the conditions III, IV, V in a slightly different form (it is not difficult to see that his II is equivalent to my IV) as a particular case of an important general result due to Hasse. He also refers, *inter alia*, to my recent paper† where I show that III, V, VI are the necessary and sufficient conditions for integer solutions of (1) subject to the restrictions typified by (2).

The first results on (1) were due to Meyer who assumes not only that a, b, c, d satisfy the conditions I, II, which it may be remarked involve no loss of generality and simplify the results, but also the further condition that no primes p_{ab} exist for which (2) holds, that is, his results applied to (1) limited only by I, II really give the solutions as restricted by my condition (4). Both his results and mine are really general in that it is very easy to transform a given equation into the required form. But the methods employed in my paper are so elementary, depending only on Legendre's criteria for the solvability of the equation

---

* *Criteria for the solution of a certain quadratic diophantine equation*, this Bulletin, vol. 37 (1931), pp. 608–614.

† *The condition for integer solutions of* $ax^2+by^2+cz^2+dt^2=0$, Journal für Mathematik, vol. 164 (1931), pp. 40–49. References to the literature are given here and also in Archibald's note.

$$Ax^2 + By^2 + Cz^2 = 0,$$

that it may be desirable to remove the restriction (4) and to deduce the condition IV instead of VI.

Let p_{ab}, p'_{ab}, etc. be the odd prime factors of $[a, b]$. Then (4) is equivalent to

$$[z, t] \not\equiv 0 \pmod{p_{ab} \text{ or } p'_{ab} \text{ etc.} \cdots},$$

that is, the 2 in (4) is really no restriction since c, d have no squared factors.

We now find the condition that (1) should be solvable subject to the six restrictions typified by (4) except that for a particular odd factor q of $[a, b]$, we are not excluding $[z, t] \equiv 0 \pmod{q}$. If we set

$$z = qz', \quad t = qt', \quad x = x', \quad y = y',$$

the equation (1) becomes

$$a'x'^2 + b'y'^2 + c'z'^2 + d't'^2 = 0, \tag{5}$$

where $a' = a/q$, $b' = b/q$, $c' = cq$, $d' = dq$. Since a', b', c', d' still satisfy I, II, we may assume that $[x', y', z', t'] = 1$ and so no three of x', y', z', t' can have a common factor. Hence x', y' are both prime to q, that is,

$$[x', y', c', d'] = 1 \text{ or } 2.$$

Hence it is clear that (5) is subject to the appropriate six restrictions typified by (4) and we can apply V and VI. But the corresponding condition V for (5) is exactly the same as V since $a'b'c'd' = abcd$ and

$$a' + b' + c' + d' = \frac{a}{q} + \frac{b}{q} + cq + dq \equiv 0 \pmod{8},$$

since $q^2 \equiv 1 \pmod{8}$.

Since $c'd' = cdq^2$, $a'b' = ab/q^2$, $a'c' = ac$, etc., the corresponding ing condition VI, on noting that $-(ab/q^2)Rq[c, d]$ means that $-(ab/q^2)Rq$ and $-(ab/q^2)R[c, d]$, becomes

$$-cdR\left[\frac{a}{q}, \frac{b}{q}\right], \quad -\frac{ab}{q^2}Rq \tag{6}$$

together with the five conditions

$$(7)\qquad \begin{matrix} -abR[c,d], & -acR[b,d], & -adR[b,c], \\ -bcR[a,d], & -bdR[a,c]. \end{matrix}$$

If $[a, b]$ is divisible by a prime p_{ab} which is not a factor of q and for which

$$(8)\qquad \left[\frac{-cd}{p_{ab}}\right] = -1,$$

(6) will not hold. Hence (6) can be satisfied only if q is divisible by every prime p_{ab} for which (2) holds. For these primes, the condition $-(ab/q^2)Rq$ becomes

$$(9)\qquad -\frac{ab}{p_{ab}^2}Rp_{ab}.$$

If there are such primes p_{ab}, VI shows that (1) is not solvable unless $[z, t] \equiv 0 \pmod{p_{ab}}$, and that it is solvable if q is the product of these p_{ab}. Hence the equation (1) has solutions subject to the five restrictions

$$(10)\qquad \begin{matrix} [x,y,c,d] = 1, 2; & [x,z,b,d] = 1, 2; & [x,t,b,c] = 1, 2; \\ [y,z,a,d] = 1, 2; & [y,t,a,c] = 1, 2; \end{matrix}$$

if and only if III and V hold and if every odd prime factor p_{ab} of $[a, b]$ for which (2) holds also satisfies (3), together with the five conditions (7).

It is clear that we can proceed now with equation (1) as restricted by (10) and remove in turn the restrictions $[x, y, c, d] = 1, 2$, etc. Thus if we now allow

$$[x, y, c, d] = r > 2,$$

and put

$$x = rx'', y = ry'', z = z'', t = t'',$$
$$a'' = ar, b'' = br, c'' = c/r, d'' = d/r,$$

then

$$a''x''^2 + b''y''^2 + c''z''^2 + d''t''^2 = 0.$$

If this has solutions subject to the five restrictions typified by (10), that is $[x'', y'', c'', d''] = 1, 2$, etc., the first condition in (7) becomes $-abr^2R(c/r, d/r)$ while the other four remain un-

changed. Hence r must include among its factors the odd prime factors p'_{cd}, of $[c, d]$, if any, for which

$$\left(\frac{-ab}{p'_{cd}}\right) = -1.$$

To the condition below (10), must be added the one arising since $[a'', b''] = r[a, b]$, that if

$$\left(\frac{-\frac{cd}{r^2}}{p'_{cd}}\right) = -1 \tag{11}$$

then also

$$\left(\frac{-abr^2/p'^2_{cd}}{p_{cd}}\right) = 1; \tag{12}$$

that is,

$$\left(\frac{-ab}{p'_{cd}}\right) = 1. \tag{13}$$

But then there is no need to take p_{cd} as a factor of r, and we must have

$$\left(\frac{-cd/p'^2_{cd}}{p'_{cd}}\right) = 1. \tag{14}$$

Hence (1) has solutions subject to the last four restrictions in (10) if and only if III and V hold together with the last four conditions in (7), and also that (3) holds when (2) holds; and (14) holds when (13) holds. We thus arrive at the condition IV when (1) is not restricted by the six conditions typified by (4).

It may be remarked that condition IV could be deduced more simply by noting the fact pointed out in my paper that my method showed that necessary and sufficient conditions for the solvability in integers not all zero of $f(x, y, z, t) = 0$, where f is an indefinite quaternary (or for that matter a ternary) quadratic form with integer coefficients and non-zero determinant, is that the congruence $f(x, y, z, t) \equiv 0 \pmod{p^n}$ should be solvable with $[x, y, z, t] \not\equiv 0 \bmod p$, for all primes $p \geqq 2$ and integers $n \geqq 0$, and that this requires a condition for only a finite number of values

of p, n. Thus for (1) we need only take $p^n = 2^4, p_{ab}^2, p_{ac}^2, \cdots$. The condition V arises from $p^n = 2^4$, and the condition IV on taking $p^n = p_{ab}^2$ and applying the result that the congruence

$$Ax^2 + By^2 + Cz^2 \equiv 0 \pmod{p^n}$$

is possible for all odd primes p if $ABC \not\equiv 0 \pmod{p}$.

Archibald's numerical example is

$$110x^2 + 770y^2 - z^2 - t^2 = 0.$$

Here only the primes p_{ab} exist and $p_{ab} = 5, 11$, since

$$[110, 770] \equiv 0 \pmod{p_{ab}}.$$

For $p_{ab} = 11$,

$$\left(\frac{-cd}{p_{ab}}\right) = \left(\frac{-1}{11}\right) = -1,$$

but

$$\left(\frac{-ab/p_{ab}}{p_{ab}}\right) = \left(\frac{-700}{11}\right) = \left(\frac{-7}{11}\right) = 1.$$

For $p_{ab} = 5$,

$$\left(\frac{-cd}{p_{ab}}\right) = 1,$$

and no other condition arises. Hence IV is satisfied. Clearly V is satisfied since

$$\tfrac{1}{2}a \equiv -1 \pmod 8, \quad \tfrac{1}{2}b \equiv 1 \pmod 8, \quad c \equiv d \pmod 8.$$

Hence the equation is solvable. Archibald gives the solution $x = 2, y = 1, z = 11, t = 33$.

THE UNIVERSITY, MANCHESTER, ENGLAND

NOTE ON AN APPLICATION OF METRIC GEOMETRY TO DETERMINANTS

BY L. M. BLUMENTHAL

1. *Introduction.* This note refers to a previous paper* published in this Bulletin. On page 754 of that paper, the following theorem is stated.

THEOREM. *If the symmetric determinant*

$$D = \begin{vmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & r_{12} & r_{13} & r_{14} \\ 1 & r_{21} & 0 & r_{23} & r_{24} \\ 1 & r_{31} & r_{32} & 0 & r_{34} \\ 1 & r_{41} & r_{42} & r_{43} & 0 \end{vmatrix}, \quad (r_{ij} = r_{ji}),$$

with $r_{ij} > 0$, $(i, j = 1, 2, 3, 4)$, $i \neq j$, *is different from zero, and the complementary minors of four of the elements in the principal diagonal vanish, then the complementary minor of the remaining element does not vanish.*

In order to prove this theorem, two cases, A and B, were considered. In Case A it was supposed that all four of the bordered complementary minors vanished. It was then shown that the fifth complementary minor (the unbordered minor that is the complementary minor of the element appearing in the first row and first column) did not vanish.

In Case B it was assumed that three of the four bordered minors and the unbordered minor, denoted by $\mathcal{E}(p_1, p_2, p_3, p_4)$, vanished. It was stated that Case B was exhausted by a study of two sub-cases, both of which were examined and found to contradict the hypotheses stated for Case B. It was concluded, then, that this case could not exist, and from this conclusion four interesting corollaries were stated. The principal interest of the paper seems to the writer to lie in these corollaries.

A communication received from W. V. Parker called my attention to the fact that the hypotheses explicitly made upon

* This Bulletin, vol. 37 (1931), pp. 752–758.

the elements of the determinant D, while sufficient for the validity of the theorem, are not sufficient to prevent the possibility of Case B occurring, and hence, the corollaries, dependent upon the impossibility of Case B for their correctness, are not true without an additional hypothesis upon D. This Parker showed by exhibiting a determinant which actually satisfied all the requirements stated, and which came under Case B.

In the present paper we give a brief proof of the theorem with the conditions as stated in the original paper, in order to clear up any doubts that might exist about the validity of the theorem itself. Further, we make explicit the additional hypothesis upon the elements of the determinant D (implicit in the previous paper) and show in some detail that with this hypothesis the two cases examined under Case B are indeed the only two types that cannot be immediately rejected. These having been proved impossible, the corollaries are seen to be valid. Finally, the last part of this note is concerned with obtaining the only class of determinants, of which the one due to Parker is a special case, for which the Case B *does* exist. It remains for me to express my thanks to him for his interest and his contribution.

2. *Proof of the Theorem.* To prove the theorem itself, we consider again the two cases A and B. The elements r_{ij} having been replaced by $(ij)^2$, where (ij) is a positive number which may be thought of as representing the distance between a point p_i and a point p_j of a quadruple p_1, p_2, p_3, p_4, the assumption in Case A that all four bordered complementary minors vanish means, as stated in the previous paper, that all four of the triples of points contained in the four points p_1, p_2, p_3, p_4 are linear. The determinant D being, by hypothesis, different from zero, the four points are not linear, and hence they form a pseudo-linear quadruple. (This is a result due to Karl Menger, explicit reference to which was given.) The determinant $\mathcal{E}(p_1, p_2, p_3, p_4)$ formed for a pseudo-linear quadruple is shown to be different from zero, and the theorem is proved for this case.

In Case B we suppose that the four complementary minors assumed to vanish consist of three of the bordered minors and the unbordered minor $\mathcal{E}(p_1, p_2, p_3, p_4)$. We must then show that the remaining minor, a bordered one, does not vanish. This is immediate, for suppose that this minor vanishes. Then

all four of the bordered minors vanish, and hence the four points p_1, p_2, p_3, p_4 are either linear or pseudo-linear. But, since the determinant $\mathcal{E}(p_1, p_2, p_3, p_4)$ is assumed to vanish, the points are not pseudo-linear, for, from Case A we have seen that the determinant $\mathcal{E}(p_1, p_2, p_3, p_4)$ formed for a pseudo-linear quadruple does not vanish. Therefore, the four points must be linear. But this is impossible, since the determinant D vanishes for a linear set of points, and, by hypothesis, the determinant D does not vanish. Hence, the assumption that the remaining minor vanishes is seen to lead to a contradiction, and the theorem is proved.

3. *The Triangular Inequality.* We now state the hypothesis that is sufficient to prevent the existence of Case B. This hypothesis is that *the elements* $r_{ij}=(ij)^2$ *of* D *are such that each of the four triples* p_1, p_2, p_3; p_1, p_2, p_4; p_1, p_3, p_4; p_2, p_3, p_4 *contained in the four points* p_1, p_2, p_3, p_4 *satisfy the triangular inequality.** This geometric way of stating the condition is made use of because of its brevity; the condition may, of course, be stated in algebraic language.

It is now to be shown that with this condition adjoined to those made explicitly in the theorem, the Case B cannot exist. Since the theorem itself has been established, we know that in the Case B, the remaining complementary minor does not vanish. This is a bordered minor, and its non-vanishing means that the triple for which it is formed is not linear. We may assume the labelling so that this non-linear triple is p_1, p_2, p_4. The other three triples being linear (since their minors vanish) we have

$$(123)(231)(312) = 0, \quad (234)(342)(423) = 0, \quad (341)(413)(134) = 0,$$

where by the symbol ijk we mean $ij+jk-ki$, and the parentheses indicate multiplication.

There are, then, 27 possible cases to examine, corresponding to the different ways in which the three bordered minors can vanish. An examination of these 27 cases yields the following results:

Twelve of the cases result in the fourth triple p_1, p_2, p_4 being linear, and hence are to be rejected, since we have seen that this

* Three points are said to satisfy the triangular inequality if their distances are such that the sum of any two is greater than or equal to the third.

triple is not linear, since the determinant formed for it does not vanish.

Nine of the cases result in the fourth triple p_1, p_2, p_4 not satisfying the triangular inequality, and are therefore to be rejected since we have supposed that all of the triples satisfy this inequality.

Two of the combinations are to be rejected since they imply that the sum of three distances equals zero.

This leaves four cases to be examined. These four cases are:

$$\text{(a)}\begin{cases}132=0,\\134=0,\\234=0,\end{cases}\quad\text{(b)}\begin{cases}123=0,\\234=0,\\341=0,\end{cases}\quad\text{(c)}\begin{cases}312=0,\\342=0,\\134=0,\end{cases}\quad\text{(d)}\begin{cases}231=0,\\423=0,\\413=0.\end{cases}$$

Case (a) is seen to reduce to sub-case alpha of Case B if we interchange the labels of p_3 and p_1. Hence, it has been shown to be impossible. Case (b) is precisely the case treated in sub-case beta of Case B, and there shown to be impossible. Case (c) is merely another labelling of case (b) in which p_1 and p_2 are interchanged. Case (d) is reduced to case (c) if the letters p_2 and p_4 are interchanged. Hence these four cases are also impossible and the Case B cannot exist. Therefore, with the additional assumption stated in this section, the corollaries of the theorem are valid.

4. *Possibility of Case* B. We now obtain a class of determinants that may occur under Case B when we drop the assumption that each of the triples satisfies the triangular inequality.

Consider the relations $312=0$; $234=0$; $413=0$. If we put $12=a$, $14=c$, $24=b$, we obtain, by substitution in the above three equations, $13=(1/2)(-a+b-c)$, $23=(1/2)(a+b-c)$, $34=(1/2)(-a+b+c)$. Since the distance (ij) is always positive, we have the positive numbers a, b, c satisfying the inequalities $a+b>c$, $b+c>a$, $b>a+c$. It will be observed that the triple p_1, p_2, p_4 does not satisfy the triangular inequality. Now, by hypothesis, the determinant $\mathcal{E}(p_1, p_2, p_3, p_4)$ vanishes. Substituting the values obtained above, and developing the determinant, we obtain

$$\mathcal{E}(p_1, p_2, p_3, p_4) = (1/16)(a+b+c)(a+b-c)$$
$$\cdot(a-b+c)(-a+b+c)^3(2ab+2ac+2bc-a^2-b^2-c^2).$$

Thus, it is seen, $\mathcal{E}(p_1, p_2, p_3, p_4)$ is zero if and only if

$$a^2 + b^2 + c^2 - 2(ab + ac + bc) = 0,$$

since the first four parentheses are seen not to vanish by virtue of the inequalities that a, b, c satisfy.

Let us calculate the value of the determinant D formed for this case. It is found that

$$D(p_1, p_2, p_3, p_4) = -\tfrac{1}{8}(-a^3 - 2b^3 - 2c^3 + 3ab^2 + 3ac^2 + 2bc^2 + 2b^2c - 6abc)^2,$$

which may be written in the form

$$D = -\tfrac{1}{8}[-(a^2+b^2+c^2-2ab-2ac-2bc)(a+2b+2c) - 16abc]^2,$$

and, on account of the condition obtained by setting $\mathcal{E}$ equal to zero, this becomes $D = -32a^2b^2c^2$.

In order to see that this constitutes the only class of determinants for which Case B exists, it is necessary and sufficient to examine each of the nine combinations referred to in §3 for which the fourth triple p_1, p_2, p_4 does not satisfy the triangular inequality. These nine combinations are

$$(1)\ \begin{cases} 312 = 0, \\ 234 = 0, \\ 413 = 0, \end{cases} \quad (2)\ \begin{cases} 123 = 0, \\ 423 = 0, \\ 134 = 0, \end{cases} \quad (3)\ \begin{cases} 231 = 0, \\ 342 = 0, \\ 341 = 0, \end{cases}$$

$$(4)\ \begin{cases} 123 = 0, \\ 342 = 0, \\ 134 = 0, \end{cases} \quad (5)\ \begin{cases} 123 = 0, \\ 234 = 0, \\ 413 = 0, \end{cases} \quad (6)\ \begin{cases} 312 = 0, \\ 423 = 0, \\ 134 = 0, \end{cases}$$

$$(7)\ \begin{cases} 312 = 0, \\ 234 = 0, \\ 341 = 0, \end{cases} \quad (8)\ \begin{cases} 231 = 0, \\ 342 = 0, \\ 413 = 0, \end{cases} \quad (9)\ \begin{cases} 231 = 0, \\ 423 = 0, \\ 341 = 0. \end{cases}$$

The first combination is the one just treated, and shown to exist. It is seen at once that the second combination may be obtained from the first by interchanging the letters p_1 and p_2; while the third is similarly obtained from the first by interchanging the labels of p_1 and p_4. Hence these two combinations are equivalent to the first and do not require further treatment.

It will be found upon examination that the last six combinations are equivalent to the fourth combination, being obtainable from this combination by suitable changes in the labelling of the points. It will be sufficient, then, to examine combination (4).

We have, writing $12=a$, $14=c$, $24=b$, and substituting in the relations (4), that

$$13 = \tfrac{1}{2}(a+b+c), \quad 23 = \tfrac{1}{2}(-a+b+c),$$
$$34 = \tfrac{1}{2}(-a-b+c),$$

the positive numbers a, b, c satisfying the inequality $c>a+b$. The determinant $\mathcal{E}(p_1, p_2, p_3, p_4)$, which vanishes, by hypothesis, takes the form, upon being developed,

$$\mathcal{E}(p_1, p_2, p_3, p_4) = (1/16)(a+b+c)^2(-a+b+c)^2(a+b-c)^2 \cdot(a^2+b^2+c^2+2ab+2bc-2ac).$$

Now evidently none of the parentheses vanishes, and hence $\mathcal{E}(p_1, p_2, p_3, p_4)$ is not zero, which contradicts the hypothesis. Hence, this case is not possible.

Thus we have seen that if we assume the elements $r_{ij}=(ij)^2$ of the determinant D to be such that each triple of points satisfies the triangular inequality, then only Case A is possible, and the four corollaries, stated at the end of the paper about which this note is written, are valid. If we do not make this assumption, then Case B can exist for only one type of combination, the one investigated in this paper. We notice that in either event, the value of the determinant is given by $-32a^2b^2c^2$, where a, b, c are positive numbers satisfying the appropriate conditions.

THE RICE INSTITUTE

ON POLYNOMIAL INTERPOLATION TO ANALYTIC FUNCTIONS WITH SINGULARITIES*

BY J. L. WALSH

Méray has given† the following illustration to show that polynomials formed from a given function by interpolation do not necessarily converge to that function. Interpolate to the function $f(z)=1/z$ by means of the polynomials $p_n(z)$ of respective degrees‡ $n=1, 2, 3, \cdots$, required to coincide with $f(z)$ in the $(n+1)$th roots of unity; this condition defines the polynomials $p_n(z)$ uniquely. Moreover, we have

$$p_n(z) \equiv z^n, \tag{1}$$

because the equation $p_n(z)=1/z$ is satisfied provided z is one of the $(n+1)$th roots of unity. Even though the sequence $p_n(z)$ is defined by interpolation from the function $f(z)=1/z$, the polynomials $p_n(z)$ do not approach the function $f(z)$ for $|z|<1$, as n becomes infinite, but approach the limit zero. It is naturally not surprising that these polynomials should fail to approach $f(z)$ for $|z|<1$, since $f(z)$ has a singularity in that region; this sequence of polynomials fails to approach the limit $f(z)$ even in a neighborhood of the curve $|z|=1$ on which interpolation takes place.

In this connection, it is worth while to recall Runge's result§ that if $f(z)$ is analytic for $|z|\leqq 1$, then the sequence of interpolating polynomials $p_n(z)$ of respective degrees n which coincide with $f(z)$ in the $(n+1)$th roots of unity converges to the limit $f(z)$ for $|z|\leqq 1$.

---

* Presented to the Society, March 25, 1932.

† Annales de l'École Normale Supérieure, (3), vol. 1 (1884), pp. 165–176. This illustration is also presented by Montel (after Méray), in his *Séries de Polynomes*, 1910, p. 51.

‡ A polynomial of the form $a_0z^n+a_1z^{n-1}+\cdots+a_n$ is said to be of degree n.

§ *Theorie und Praxis der Reihen*, 1904, p. 137. This method of Runge's has been more systematically developed by Fejér, Göttinger Nachrichten, 1918, pp. 319–331, and by L. Kalmár, *Mathematikai és Physikai Lapok*, 1926, pp. 120–149, but only for interpolation in points on the boundary of a region of a function known to be analytic in that region.

Méray's illustration is of such simplicity, directness, and beauty, and apparently has stood alone for so long, that it seems worth while to furnish it with some companionship. The purpose of the present note is to provide such companionship by the proof of the following theorem.

THEOREM 1. *Let $f(z)$ be an arbitrary function continuous for $|z|=1$. Let the polynomials $p_n(z)$ of respective degrees n, $(n=1, 2, \cdots)$, be defined by the requirement of coinciding with $f(z)$ in the $(n+1)$th roots of unity. Then the sequence $p_n(z)$ approaches the limit*

$$f_1(z) = \frac{1}{2\pi i}\int_{|t|=1} \frac{f(t)dt}{t-z} \tag{2}$$

for $|z|<1$, uniformly for $|z| \leqq r<1$.

Lagrange's interpolation formula for the polynomial $p_n(z)$ of degree n which takes on the values $K_1, K_2, \cdots, K_{n+1}$ at the $n+1$ distinct points $z_1, z_2, \cdots, z_{n+1}$ is

$$p_n(z) = \sum_{k=1}^{n+1} \frac{K_k}{p'(z_k)} \frac{p(z)}{z-z_k},$$

where $p(z)=(z-z_1)(z-z_2)\cdots(z-z_{n+1})$; the polynomial $p_n(z)$ is uniquely determined by these requirements. Under the circumstances of the theorem, we set $z_k=\omega^k$, where $\omega=e^{2\pi i/(n+1)}$, $p(z)=z^{n+1}-1$. It follows that we have

$$p_n(z) = \sum_{k=1}^{n+1} f(\omega^k) \frac{\omega^k(z^{n+1}-1)}{(n+1)(z-\omega^k)}. \tag{3}$$

With the exception of the term z^{n+1} in the numerator, which approaches zero, equation (3) suggests computation of the integral which appears in (2) by division of the circle C defined by $|z|=1$, at the points ω^k. We have

$$f_1(z) = \lim_{n\to\infty} \frac{1}{2\pi i} \sum_{k=1}^{n+1} \frac{f(\omega^k)(\omega^{k+1}-\omega^k)}{\omega^k - z}, \tag{4}$$

$$\lim_{n\to\infty} [f_1(z) - p_n(z)]$$

$$= \lim_{n\to\infty} \left[\frac{1}{2\pi i} + \frac{z^{n+1}-1}{(n+1)(\omega-1)}\right] \sum_{k=1}^{n+1} \frac{\omega^k(\omega-1)f(\omega^k)}{\omega^k - z}. \tag{5}$$

By (4), the summation on the right-hand side approaches the limit $2\pi i f_1(z)$, which is continuous for $|z|<1$, and the limit is approached uniformly for $|z|\leqq r<1$.* The quantity $(n+1)(\omega-1)$ approaches as its limit $2\pi i$, for we have

$$\omega = \cos\frac{2\pi}{n+1} + i\sin\frac{2\pi}{n+1},$$

$$\frac{(n+1)(\omega-1)}{2\pi i} = \frac{\cos\dfrac{2\pi}{n+1}-1}{\dfrac{2\pi i}{n+1}} + \frac{\sin\dfrac{2\pi}{n+1}}{\dfrac{2\pi}{n+1}},$$

which approaches the limit unity. The square bracket in the right-hand member of (5) thus approaches zero for $|z|<1$, uniformly for $|z|\leqq r<1$, and the factor of this bracket is bounded uniformly in z and n, $|z|\leqq r<1$, so the proof of our theorem is complete.

The theorem and proof are obviously valid if the given function $f(z)$ is not supposed continuous on C: $|z|=1$, but merely integrable in the sense of Riemann.

The sequence $p_n(z)$ clearly converges uniformly in a region containing the curve C in its interior when and only when the given function $f(z)$ or its analytic extension is analytic on and within C. The first part of this result was proved by Féjer (loc. cit.), or compare Walsh.† Reciprocally, if the sequence $p_n(z)$ converges uniformly for $|z|\leqq\rho>1$, we shall prove $f(z)$ analytic on and within C. The limit of the sequence $p_n(z)$ is analytic for $|z|<\rho$. The obvious equation $p_{m(n+1)}(\omega^k)=f(\omega^k)$, $(m=1, 2, \cdots)$, where $\omega=e^{2\pi i/(n+1)}$, implies the convergence of the sequence $p_n(z)$ to the function $f(z)$ at each point ω^k, hence at a set of points everywhere dense on C. The continuity of $f(z)$ for $|z|=1$ and the analytic character of the limit of the sequence $p_n(z)$ for $|z|<\rho$ then implies the identity on C of those two functions, hence the fact that $f(z)$ or its analytic extension is analytic for $|z|<\rho$. If the sequence $p_n(z)$ converges uniformly for $|z|\leqq 1$, the function $f(z)$ is analytic for $|z|<1$, and continuous for $|z|\leqq 1$.

* Compare Runge, *Acta Mathematica*, vol. 6 (1885), pp. 229–244; Montel, loc. cit., p. 57; or Osgood, *Funktionentheorie*, 1928, pp. 579–581.

† Transactions of this Society, vol. 34 (1932), pp. 22–74, §11.

The following result is a not uninteresting complement to the main theorem already proved.

THEOREM 2. *The function $f_1(z)$ which is the limit for $|z|<1$ of the sequence $p_n(z)$ of the interpolating polynomials for $f(z)$ is also the limit for $|z|<1$, uniformly for $|z|\leqq r<1$, of the sequence of polynomials $P_n(z)$ of respective degrees $n=1, 2, \cdots$, of best approximation to $f(z)$ on C in the sense of least squares.*

The polynomial $P_n(z)$ is that polynomial of degree n for which

$$\int_C |f(z)-P_n(z)|^2\,|dz|$$

is least. Such a polynomial is known to exist and be unique; it is defined* by

$$P_n(z)=c_0+c_1z+\cdots+c_nz^n, \tag{6}$$

$$c_n=\frac{1}{2\pi}\int_C f(z)\bar{z}^n\,|dz|=\frac{1}{2\pi i}\int_C f(z)\,\frac{dz}{z^{n+1}}.$$

We have the relation

$$f_1(z)=\frac{1}{2\pi i}\int_C\frac{f(t)dt}{t-z}=\frac{1}{2\pi i}\int_C f(t)dt\left[\frac{1}{t}+\frac{z}{t^2}+\frac{z^2}{t^3}+\cdots\right].$$

The infinite series converges uniformly in z and t for t on C, $|z|\leqq r<1$, and hence may be integrated term by term. Thus we have from (6) $f_1(z)=c_0+c_1z+c_2z^2+\cdots$, a series which is uniformly convergent for $|z|\leqq r<1$; this includes the relation to be proved:

$$f_1(z)=\lim_{n\to\infty}P_n(z),\qquad |z|<1.$$

The polynomial $P_n(z)$ is the sum of the first $n+1$ terms of the Maclaurin development of $f_1(z)$.

Several particular cases are worth mentioning. If there exists a function analytic for $|z|<1$, continuous for $|z|\leqq 1$, which coincides with the given function $f(z)$ on C, then $f_1(z)$ naturally coincides with this function. If there exists a function $F(z)$ analytic for $|z|>1$ (including the point $z=\infty$), continuous for $|z|\geqq 1$, which coincides with the given function $f(z)$ on C, then each $c_n(n>0)$ vanishes, and we have

* See for instance Kowalewski, *Determinantentheorie*, 1909, §137.

$$\lim_{n\to\infty} p_n(z) = \lim_{n\to\infty} P_n(z) = c_0 = F(\infty), \quad |z| < 1;$$

here is included the case $f(z) = 1/z^k$, k a positive integer ($k=1$ is Méray's example), and we have $f_1(z) \equiv 0$. More generally, if on C the function $f(z)$ can be expressed as $f(z) = f_1(z) + f_2(z)$, where $f_1(z)$ is analytic for $|z| < 1$, continuous for $|z| \leqq 1$, and where $f_2(z)$ is analytic for $|z| > 1$ including the point at infinity, is continuous for $|z| \geqq 1$, and vanishes at infinity, then we have

$$\lim_{n\to\infty} p_n(z) = \lim_{n\to\infty} P_n(z) = f_1(z), \quad |z| < 1,$$

uniformly for $|z| \leqq r < 1$. If this function $f_1(z)$ is analytic for $|z| < \rho > 1$, then we have

$$\lim_{n\to\infty} P_n(z) = f_1(z), \quad |z| < \rho, \text{ uniformly for } |z| \leqq \rho' < \rho,$$

no matter what $f_2(z)$ may be, but we have

$$\lim_{n\to\infty} p_n(z) = f_1(z), \quad |z| < \rho, \text{ uniformly for } |z| \leqq \rho' < \rho,$$

if and only if $f_2(z)$ vanishes identically.

Another illustration of the close connection between interpolation in the $(n+1)$th roots of unity and approximation on C in the sense of least squares has recently been indicated by the present writer (loc. cit.).

If the function $f(z)$ is analytic for $|z| < T > 1$, then for the polynomials $p_n(z)$ and $P_n(z)$ already defined we have

$$\lim_{n\to\infty} [p_n(z) - P_n(z)] = 0, \quad |z| < T^2, \textit{ uniformly for } |z| \leqq R < T^2$$

even though $f(z)$ has singularities for $T \leqq |z| < T^2$.

It is not to be supposed (compare Kalmár, loc. cit.) that interpolation in points *arbitrarily* chosen on C is always equivalent in the sense illustrated to approximation on C in the sense of least squares, even when the points of interpolation $z_n^{(k)}$ for the interpolating polynomial $p_n(z)$ of degree n are such that the limit of the maximum distance between successive points $z_n^{(k)}$ on C for a given n approaches zero with $1/n$. We give an example to illustrate this fact, where the $n+1$ points of interpolation for the polynomial $p_n(z)$ are the roots of

$$\left(\frac{1-\alpha z}{\alpha - z}\right)^{n+1} = 1, \quad \alpha > 1. \tag{7}$$

These points of interpolation are thus the transforms in the z-plane of the $(n+1)$th roots of unity in the w-plane, under the transformation

$$w = \frac{1-\alpha z}{\alpha - z}, \tag{8}$$

which leaves C invariant and transforms the interior and exterior of C respectively into the interior and exterior of C. It is simpler to study the situation in the w-plane. We interpolate to the function $f(z) = 1/(w+T)$, $0 < T < 1$, in the points w, $w^{n+1} = 1$, by rational functions $p_n(z) = F_n(w)$ of respective degrees (in z or w) n whose poles coincide in the points $w = \alpha$. The reader can easily verify the formula

$$\begin{aligned} p_n(z) &= F_n(w) \\ &= \frac{(T+\alpha)^n(w^{n+1}-1) + [(-1)^n + T^{n+1}](w-\alpha)^n}{[(-1)^n + T^{n+1}](w+T)(w-\alpha)^n}; \end{aligned} \tag{9}$$

the expression in terms of z is found from (8). It appears from (9) that one may write

$$p_n(z) = \frac{1}{w+T} + \frac{(T+\alpha)^n(w^{n+1}-1)}{[(-1)^n + T^{n+1}](w+T)(w-\alpha)^n},$$

so that, for $|w| < 1$, we have actual convergence when and only when the condition

$$T + \alpha < |w - \alpha| \tag{10}$$

is satisfied. Condition (10) is equivalent to the condition that w should lie exterior to a certain circle whose center is α and which cuts $C: |w| = 1$. Thus the sequence $p_n(z)$ converges in only a part of the unit circle $|z| \leqq 1$, and in that part converges to the original function $1/(w+T)$.

It would be of interest to extend the main theorem of this note to the study of curves other than the unit circle; compare the references already given to Fejér and Kalmár.

HARVARD UNIVERSITY

ANALYTIC STUDY OF RATIONAL QUINTIC SURFACES HAVING NO MULTIPLE CURVES

BY H. N. HUBBS

1. *Introduction.* The purpose of this paper is to derive the equations of certain of the rational quintic surfaces without multiple curves discussed synthetically by Montesano.* The equations of the surfaces are found by applying Cremona transformations to certain well known rational surfaces of order three or four.

2. *Surface of Order Five with Four Triple Points.* This surface is the transform by the cubic transformation T_{tet}† of a general cubic surface ϕ_3 through the vertices of the tetrahedron. The equation of the surface is

$$\begin{aligned}\phi_5 \equiv\ & y_1^2[y_2^2u + y_3^2u' + y_4^2u'' + Ay_2y_3y_4] \\ & + y_1[y_4^2\phi_2 + y_2y_3(By_3y_4 + y_2u''')] \\ & + y_2y_3y_4[Cy_3y_4 + Dy_2y_4 + Ey_2y_3] = 0,\end{aligned}$$

where u, u', u'', u''' are linear in (y_3, y_4), (y_2, y_4), (y_2, y_3), (y_3, y_4), respectively, and ϕ_2 is quadratic in (y_2, y_3), and where A, B, C, D and E are constants. The points whose coordinates are (1,0,0,0), (0,1,0,0), (0,0,1,0), (0,0,0,1) are triple points, with non-composite tangent cones at each of the points.

3. *The Surface ϕ_5 with Three Ordinary Triple Points and a Tacnode.* The surface ϕ_5 is the transform by T_{tet} of a quartic surface with a double conic passing through three of the vertices of the tetrahedron and having the fourth vertex at a general point of the surface. The section of ϕ_5 by a plane through the tacnode and two triple points is a straight line and a pair of conics passing through these points.

The equation of ϕ_4 with a double conic is

$$\phi_4 \equiv [\ \textstyle\sum a_ix_jx_k]^2 - 4x_4^2[\psi_2 + x_4\psi_1] = 0,$$
$$(i, j, k = 1, 2, 3, 4, \cdots, i \neq j \neq k),$$

* Montesano, Napoli Rendiconti, (3), vol. 7 (1901), pp. 67–106.

† Hudson, *Cremona Transformations in Plane and Space*, Cambridge University Press, 1927, pp. 301–303.

where ψ_1 and ψ_2 are linear and quadratic, respectively, in the variables (x_1, x_2, x_3). The surface ψ_5 is

$$\phi_5 \equiv y_4^3\phi_1^2 - 4y_4\psi_4 - 4y_1y_2y_3\phi_2 = 0,$$

where ϕ_1 is linear in (y_1, y_2, y_3), ψ_4 is quadratic, and ϕ_2 linear in (y_2y_3, y_1y_3, y_1y_2). The point (0, 0, 0, 1) is a tacnode on ϕ_5, and (0, 0, 1, 0), (0, 1, 0, 0), (1, 0, 0, 0) are triple points with non-composite tangent cones.

4. *The Surface ϕ_5 with Three Ordinary Triple Points and a Tacnode, the Tacnode lying with one of the Triple Points on a Line of ϕ_5 Situated in the Tangent Plane at the Tacnode.* This surface is the transform by T_{tet} of a rational surface ϕ_4 of order four, having a double line and two double points in a plane through the double line. One fundamental point of the transformation is on the double line, two at the double points, and the fourth a general point of ϕ_4. The plane of the double points and double line is tangent to ϕ_4 along the line joining the double points. The transform of ϕ_4 by T_{tet} is

$$\phi_5 \equiv (y_2 - y_1)^2y_4^3 + (y_2 - y_1)^2y_4\psi_2 + y_4\psi_4 + \psi_5 = 0,$$

where ψ_2, ψ_4, and ψ_5 are forms in (y_1, y_2, y_3) of the order of their subscripts. The triple points are then (1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0), and the tacnode is (0, 0, 0, 1).

5. *The Surface ϕ_5 with Two Triple Points and Two Tacnodes.* A space Cremona transformation of order three is defined by the web of surfaces of order three passing through three fixed conics k_2, k_2', k_2'' which lie in distinct planes, have one point in common, and meet by pairs in three points. The conjugate system is defined by cubic surfaces having in common three non-concurrent coplanar lines and a space cubic curve meeting each line once.

The inverse of this transformation carries a quadric surface through one of the lines and the intersection of the other two into a surface ψ_5 of order five having two of the fundamental conics as double curves and the third a simple curve. The transformation T_{tet} with fundamental points at the intersections of these conics carries ψ_5 into a surface ϕ_5 which has two triple points and two tacnodes. The tacnodes are the images of the double conics of ψ_5.

The equation ϕ_5 obtained by the above procedure is

$$\begin{aligned}\phi_5 \equiv\ & y_4^2[y_2y_3u_1 + Ay_1y_3u_2 + By_1y_2u_3 + Cy_1y_2y_3 \\ & + u_1(Dy_3u_2 + Dy_2u_3 + Ey_2y_3)] \\ & + y_4[y_2y_3u_4 + Ay_1y_3u_5 + By_1y_2u_6 + u_1(Fy_3u_5 + Dy_2u_6) \\ & + u_4(Fy_3u_2 + Dy_2u_3 + Ey_2y_3)] + u_4u_5u_6 = 0,\end{aligned}$$

where

$$u_1 \equiv b_4y_2 + \frac{b_3c_4}{c_2}y_3, \qquad u_4 \equiv y_2(b_1y_1 + b_3y_3) + \frac{b_3c_1}{c_2}y_1y_3,$$

$$u_2 \equiv a_4y_1 + \frac{a_3c_4}{c_1}y_3, \qquad u_5 \equiv y_1(a_2y_2 + a_3y_3) + \frac{a_3c_2}{c_1}y_2y_3,$$

$$u_3 \equiv a_4y_1 + \frac{a_2b_4}{b_1}y_2, \qquad u_6 \equiv y_1(a_2y_2 + a_3y_3) + \frac{a_2b_3}{b_1}y_2y_3;$$

and the double conics of the first transformation are

$$x_1 = 0, \quad \sum a_ix_jx_k = 0, \quad (i, j, k = 2, 3, 4, \cdots, i \neq j \neq k),$$

$$x_2 = 0, \quad \sum b_ix_jx_k = 0, \quad (i, j, k = 1, 3, 4, \cdots, i \neq j \neq k),$$

$$x_3 = 0, \quad \sum c_ix_jx_k = 0, \quad (i, j, k = 1, 2, 4, \cdots, i \neq j \neq k).$$

The points (0, 1, 0, 0) and (0, 0, 1, 0) are tacnodes and the points (1, 0, 0, 0), (0, 0, 0, 1) are triple points. The line $y_2 = 0$, $y_3 = 0$ lies on the surface.

6. *The Surface ϕ_5 with One Triple Point and Three Tacnodes.* The inverse of the first transformation of §5 carries a quadric surface through the intersections of the fundamental straight lines into a surface ψ_6 of order six having the fundamental conics as double curves, their common intersection a four-fold point, and the three points of intersection of these conics by pairs three-fold points. The transformation T_{tet}, with fundamental points at these multiple points, carries ψ_6 into a ϕ_5 with one triple point and three tacnodes. The procedure described above gives

$$\begin{aligned}\phi_5 \equiv\ & y_4^2[y_1y_2y_3 + u_1(y_3u_2 + y_2u_3 + y_2y_3) \\ & + y_1u_2(u_3 + y_3) + y_1y_2u_3] + y_4[u_1(y_3u_5 + y_2u_6) \\ & + u_4(y_3u_2 + y_2u_3 + y_2y_3) + y_1u_2u_6 + y_1u_5(u_3 + y_3) \\ & + y_1y_2u_6] + u_4(y_3u_5 + y_2u_6) + y_1u_5u_6 = 0,\end{aligned}$$

where u_i are as indicated in §5. The point (0, 0, 0, 1) is an ordinary triple point and the points (1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0) are tacnodes. The tangent cones at the tacnodes are, respectively,

$$(a_2y_2 + a_3y_3 + a_4y_4)^2 = 0, \quad (b_1y_1 + b_3y_3 + b_4y_4)^2 = 0,$$
$$(c_1y_1 + c_2y_2 + c_4y_4)^2 = 0.$$

7. *The Surface ϕ_5 with an Ordinary Triple Point and a Tacnodal Triple Point.* A Cremona transformation is defined by the web of quadric surfaces containing a fixed conic and an arbitrary point. A special case of this transformation arises when the point is on the fixed conic.

If the conic of this web is tangent to a rational quartic surface with a double line at a general point of the line, the transformation defined by the web carries the quartic surface into a quintic having an ordinary triple point and a triple point with an adjacent infinitesimal double line,* or tacnodal triple point, both lying on the fundamental conic.

The equation of the quintic obtained by the above transformation is

$$\phi_5 \equiv y_4^2 y_3^2 \psi_1(y_1, y_2) + y_4y_3\phi_1(y_1, y_2, y_3)\cdot\phi_2(y_1, y_2) + y_3^2 y_1\psi_2(y_1, y_2)$$
$$+ y_3\psi_4(y_1, y_2) + \psi_5(y_1, y_2) = 0.$$

The point (0, 0, 1, 0) is an ordinary triple point and (0, 0, 0, 1) a tacnodal triple point. In the plane $y_3 = 0$ are five straight lines on the surface, images of the residual intersections of the fundamental conic and the quartic surface. On ϕ_5 are eighteen conics passing through the triple points and lying by pairs in nine planes through the triple points.

8. *The Surface ϕ_5 with One Tacnodal Triple Point and One Tacnode.* In the following the vertices of the tetrahedron of reference are the points A_i, those of the conjugate system B_i, $(i = 1, 2, 3, 4)$.

A rational surface of order four of the third type of Noether has a double point A_4 which is a cusp in a general plane section through it. The surface has a simple line passing through the double point; a general section through this line is a cubic curve

* Segre, Annali di Matematica, (2), vol. 25 (1896), pp. 1–53.

having the line as inflectional tangent at A_4. The tangent cone at A_4 is the plane p taken twice; this is tangent to ϕ_4 along the line. A section by this plane is a conic tangent to the line at A_4.

In the transformation of §7, let the conic be tangent to ϕ_4 at A_4. $T \equiv [F_2 \equiv C_2 A_4 p,\ F_2' \equiv C_2' B_4 p']$. Under $T, \phi_4 \sim \phi_5 :: B_3^2 B_4^3$, where B_3 is a tacnode, and B_4 is a tacnodal triple point. The conic C_2' is on ϕ_5. As in (7), the point B_4 is a tacnodal triple point, the tangent planes being p' taken twice, and p_1' containing C_2'.

The equation of the quintic surface obtained by this method is

$$\begin{aligned}\phi_5 \equiv\ & y_1 y_3^2 y_4^2 + 2y_3 y_4 [y_1 y_3 \phi_1 + \phi_3 - y_1(y_1 y_2 + y_1^2 + y_2^2)] \\ & + (y_1 y_2 + y_1^2 + y_2^2)[y_1(y_1 y_2 + y_1^2 + y_2^2) - 2(y_1 y_3 \phi_1 + \phi_3)] \\ & - y_1^2 y_3^3 + y_1 y_3^2 C_2 + y_1 y_3 C_3 + y_1 C_4 = 0,\end{aligned}$$

where ϕ_1, ϕ_3, and C_i are forms in (y_1, y_2) of the order of their subscripts.

The point (0, 0, 0, 1) is a tacnodal triple point and (0, 0, 1, 0) is a tacnode lying on a line of ϕ_5 situated in the simple tangent plane at the triple point. The section of ϕ_5 by the simple tangent plane at the triple point is the conic $y_3 y_4 - y_2^2 = 0$ and the line $y_2 = 0$ counted three times; this line is the image of the simple line of ϕ_4.

9. *The Surface ϕ_5 with an Ordinary Triple Point and an Oscnode.* A quadratic Cremona transformation is defined by the quadric surfaces F_2 having in common two generators and osculating at their point of intersection.* Let the generators be l_1 and l_2 and their point of intersection A_1. The transformation is

$$T \equiv [F_2 :: l_1, l_2, A_1, F_2' \equiv l_1', l_2', B_1].$$

Let A_1 be a generic point on a monoidal quartic surface ϕ_4, and let l_1, l_2 each osculate ϕ_4 at A_1. Let the triple point of ϕ_4 be A_4. Under $T, \phi_4 \sim \phi_5$ with an ordinary triple point at B_4 and an oscnode at B_1.

A general straight line through A_1 has three residual intersections with ϕ_4; hence the image straight line has three intersections with ϕ_5 not at B_1. A general straight line meets ϕ_4 in

* Hudson, loc. cit., pp. 197–198.

four points; hence its image conic has four points in common with ϕ_5 not at B_1; that is, at B_1 are three consecutive double points on the image conic or B_1 is an oscnode on ϕ_5. The equation of ϕ_4 is

$$\phi_4 \equiv \psi_3 x_4 + \psi_4 = 0,$$

where

$$\psi_3 \equiv A x_1^3 + w_1 x_1^2 + w_2 x_1 + w_3,$$
$$\psi_4 \equiv B x_1^2 x_2 x_3 + u_2 x_1 + u_4,$$

and where w_i, u_i are binary forms in (x_2, x_3) of order i. The line $x_2 = 0$, $x_4 = 0$ osculates ϕ_4 at (1, 0, 0, 0). The equations of transformation are

$$\rho x_1 = B(y_1 y_4 - y_2 y_3),\ \rho x_2 = A y_2 y_4,\ \rho x_3 = y_3 y_4,\ \rho x_4 = y_4^2.$$

The equation of the resulting quintic is

$$B^2(y_1 y_4 - y_2 y_3)^2[ABy_1 + w_1] + B(y_1 y_4 - y_2 y_3)[w_2 y_4 + u_3] + y_4[w_3 y_4 + u_4] = 0,$$

where w_i, u_i, are the above forms in (y_2, y_3). The point (0, 0, 0, 1) is a triple point and (1, 0, 0, 0) is an oscnode.

10. *The Surface ϕ_5 with an Oscnode and a Tacnode.* Applying the transformation of §9 to a quartic surface of the first type of Noether, we find

$$\phi_5 \equiv B^2(y_1 y_4 - y_2 y_3)^2[ABy_1 + y_4 + u_1] + B(y_1 y_4 - y_2 y_3)[u_2 y_4 + u_3] + u_4 y_4 = 0,$$

where u_i are binary forms in the variables (y_2, y_3) of order i. The point (0, 0, 0, 1) is a tacnode and (1, 0, 0, 0) an oscnode.

11. *The Surface ϕ_5 with an Oscnode and a Double Point of the First Order.* Such a surface is the transform by the above transformation of a quartic of the second type of Noether. The equation of this quintic is

$$\begin{aligned}\phi_5 \equiv\ & \phi_1^2 y_4^3 + 2y_4^2[y_3\phi_1(y_3 + \psi_1) + \phi_3] \\ & + y_4\{y_3^4 + 2y_3^3\psi_1 - [D\psi_1 + F\phi_1 + D(2Fy_1 + Gy_2) \\ & + BK_1 y_1 + K_2 y_2]y_2 y_3^2 + y_2 y_3\phi_2 + y_2^3\phi_1'\} \\ & + y_2 y_3\psi_2 + AB^2 y_1 y_2^2 y_3^2 = 0,\end{aligned}$$

where

$$D=b_1B, \quad E=b_2A, \quad F=c_1B, \quad G=c_2A, \quad \phi_1=Dy_1+Ey_2,$$

and where

$$\psi_1 = Fy_1 + (G - D)y_2,$$
$$\phi_3 = AB^3y_1^3 + B^2K_3y_1^2y_2 + BK_4y_1y_2^2 + K_5y_2^3,$$
$$\phi_2 = -2AB^3y_1^2 + B(K_6 - 2BK)y_1y_2 - BK_4y_2^2,$$
$$\phi_1' = BK_7y_1 + (K_8 - BK_4)y_3,$$
$$\psi_3 = (2DF - BK_1)y_2y_3^2 + B(BK_3 - K_6)y_2^2y_3 - 2Fy_3^3 - BK_7y_2^3.$$

The point (1, 0, 0, 0) is an oscnode, and (0, 0, 0, 1) is a double point of the first order.

12. *The Surface ϕ_5 with an Oscnode and a Double Point of the Second Order.* Such a surface is the transform of a quartic surface of the third type of Noether. The equation of this quintic is

$$\phi_5 \equiv B^2(y_1y_4 - y_2y_3)^2\phi_1 + B(y_1y_4 - y_2y_3)[\phi_3 + y_2y_4\psi_1] + y_2^2y_4[Ay_2y_4 + \phi_2] = 0,$$

where ϕ_1 is linear in all variables, and where ψ_1, ϕ_2, ϕ_3 are forms in (y_2, y_3) of the order of their subscripts. The point (1, 0, 0, 0) is an oscnode, and (0, 0, 0, 1) is a double point of the second order.

13. *The Surface ϕ_5 as a General Member of a Homaloidal Family of a Cremona Space Transformation.* The quintic surfaces discussed above are all rational and their (1, 1) representations on a plane π are known. Cremona* has shown that the set of space transformations having ϕ as a general member of the first homaloidal family corresponds to a set of plane Cremona transformations in π.

If, by a Cremona transformation, ϕ_n is the transform of a rational ψ_m which is a general member of the homaloidal families of k Cremona transformations, then ϕ_n will serve as the general member of the homaloidal families of k Cremona transformations.

---

* Cremona, Istituto Lombardo Rendiconti, (2), vol. 4 (1871), pp. 269–279; and Annali di Matematica, (2), vol. 5 (1871), pp. 131–162.

14. *Conclusion.* Each of the rational quintic surfaces discussed above can serve as a general member of the homaloidal family of a Cremona transformation. The surfaces of §§5, 6, 9 are the transforms, by Cremona transformations, of a general quadric surface. The surface of §2 is the transform of a general ϕ_3; that of §10 is the transform of a quartic surface of the first type of Noether which is the transform of a general ϕ_3. The surfaces of §§8, 12 are the transforms of a quartic surface of the third type of Noether which is the transform of a special quartic of the first type of Noether with a double point in the tangent plane at the tacnode; this plane is tangent to the surface along the line joining the double points,* and this surface is a transform of a general ϕ_3. The surfaces of §§7, 11 are transforms of a quartic with a double line.† The surface of §4 is the transform of a quartic with a double line and two double points coplanar with the double line; the homaloidal family of which this ϕ_4 is a general member have in common the double line and also have contact along the line joining the double points. The surface of §3 is the transform of a quartic with a double conic.‡

CORNELL UNIVERSITY

* Noether, Mathematische Annalen, (3), vol. 33 (1889), pp. 546–571.
† Montesano, Roma Rendiconti, (4), vol. 5–2 (1889), pp. 123–130.
‡ Aroldi, Giornale di Matematiche, (3), vol. 11 (1920), pp. 175–192.

ADJOINT SYSTEMS IN THE PROBLEM OF MAYER UNDER GENERAL END-CONDITIONS*

BY S. B. MYERS

1. *Introduction.* Necessary conditions in the problems of Lagrange and Mayer with variable end points, analogous to the Euler equations and transversality condition in the ordinary problem of the calculus of variations in the plane, have recently been given by Morse and the author.† The methods used in the derivation of these conditions, as well as those used by Bliss‡ in his derivation of somewhat more complicated transversality conditions in the same problem, consist primarily of certain proofs which more properly belong to the theory of differential equations.

In the present paper, certain adjoint relationships in the problem hitherto unnoticed are pointed out which make it possible to derive the necessary conditions mentioned above merely by a consideration of the compatibility of the adjoint system of a certain set of differential equations and boundary conditions. In fact, in the anormal case it is shown that these necessary conditions form an adjoint system of the differential and terminal equations of variation. The normal case is treated essentially by making it the abnormal case of a slightly different problem. Such considerations have the advantage of referring a major part of the proofs to the theory of differential equations, and thus greatly simplifying them.

2. *The Definition of the Adjoint System.* Consider the set of m linear homogeneous differential equations of the first order§ in the n variables $\eta_1, \cdots, \eta_n$,

$$L_\beta(\eta) \equiv p_{\beta i}\eta_i' + q_{\beta i}\eta_i = 0, \quad m \leqq n, \tag{1}$$

$$(\beta = 1, \cdots, m;\ i = 1, \cdots, n),$$

* Presented to the Society, April 3, 1931.

† Morse and Myers, Proceedings of the American Academy of Arts and Sciences, vol. 66 (1931), pp. 235–253.

‡ Bliss, Transactions of this Society, vol. 19 (1918), pp. 305–314.

§ The usual convention of summation with respect to an index repeated in the same term is used throughout.

where the functions $p_{\beta i}(x)$, $q_{\beta i}(x)$ are continuous on the interval $x^1 \leqq x \leqq x^2$ and the matrix $\|p_{\beta i}(x)\|$ is of rank m throughout this interval.

By the *adjoint set of equations* we shall mean the following set of n equations in the m dependent variables $\lambda_1, \cdots, \lambda_m$:

$$M_i(\lambda) \equiv -\frac{d}{dx}(p_{\beta i}\lambda_\beta) + q_{\beta i}\lambda_\beta = 0, \tag{2}$$

$$(\beta = 1, \cdots, m;\ i = 1, \cdots, n).$$

By a *continuous solution* of these quasi-differential equations* is meant a set (λ) of m continuous functions of x such that the functions $p_{\beta i}\lambda_\beta$ are of class C' and the equations are satisfied.

We note the following identity† which holds for arbitrary functions $(\eta) = (\eta_1, \cdots, \eta_n)$ of class C' and arbitrary continuous functions $(\lambda) = (\lambda_1, \cdots, \lambda_m)$ such that the functions $p_{\beta i}\lambda_\beta$ are of class C':

$$\int_{x^1}^{x^2} [\lambda_\beta L_\beta(\eta) - \eta_i M_i(\lambda)]\,dx \equiv [p^s_{\beta i}\lambda^s_\beta \eta^s_i]^2_1. \tag{3}$$

Here η^s_i represents the value of $\eta_i(x)$ at $x = x^s$, $(s = 1, 2)$, and λ^s_β the value of $\lambda_\beta(x)$ at $x = x^s$.

Consider any set of homogeneous linear boundary conditions on η^s_i. By a set of *adjoint boundary conditions* of these conditions shall be meant a set of linear homogeneous boundary conditions on λ^s_β with the following properties.

I. The bilinear form $[p^s_{\beta i}\lambda^s_\beta \eta^s_i]^2_1$ vanishes for all λ^s_β which are subject to these conditions while the η^s_i are subject to the original boundary conditions.

II. Of all sets of conditions with the above property (I), these conditions restrict λ^s_β as little as possible; that is, the points in $2m$-space with coordinates λ^s_β fill out a subspace of maximum possible dimensionality.

The original boundary conditions may be written in the form

$$\eta^s_i = c^s_{ih}u_h, \quad (h = 1, \cdots, r;\ s = 1, 2;\ i = 1, \cdots, n), \tag{4}$$

where $(u) = (u_1, \cdots, u_r)$ are parameters which are allowed to

* See Bôcher, *Leçons sur les Méthodes de Sturm*, p. 10.

† The notation $[\quad]^2_1$ means the value of $[\quad]$ for $s = 2$ minus its value for $s = 1$.

take on any values whatever, and $\|c^s_{ih}\|$ is a $(2nr)$-dimensional matrix of constants.

THEOREM 1. *A necessary and sufficient condition that a set of homogeneous linear boundary conditions on λ^s_β in the form*

$$b^s_{\beta\delta}\lambda^s_\beta = 0, \quad (\delta = 1, \cdots, p;\ p \text{ arbitrary}), \tag{5}$$

be a set of adjoint boundary conditions of (4) *is that they be equivalent to the following conditions*:

$$V_h \equiv [p^s_{\beta i}\lambda^s_\beta c^s_{ih}]^2_1 = 0, \qquad (h = 1, \cdots, r). \tag{6}$$

The proof is simple, and will be left to the reader.

Equations (2) and conditions (6) together form an *adjoint system* of (1) and (4).

3. *Compatibility of the Adjoint System.* In applying our definition of an adjoint system to the problem of Mayer, the chief interest will lie in the question as to when an adjoint system has a solution. To answer this question, we start by adjoining to the differential equations (1) $n-m$ more differential equations so that the extended set of equations becomes

$$p_{ki}\eta_i' + q_{ki}\eta_i = \zeta_k(x), \tag{7}$$

$$(\zeta_\beta \equiv 0), (\beta = 1, \cdots, m;\ i, k = 1, \cdots, n).$$

Here the functions $q_{\tau i}(x)$, $p_{\tau i}(x)$, $(\tau = m+1, \cdots, n)$, are any continuous functions subject only to the restriction that the determinant $|p_{ki}(x)|$ be different from zero along (x^1, x^2).*

If we have any solution (η) of (1) of class C', we can use (7) to define continuous functions $\zeta_\tau(x)$ which we will say *correspond* to (η). On the other hand, we can choose the functions ζ_τ as arbitrary continuous functions of x and obtain a solution (η) of (7) of class C' taking on arbitrary values at one end point. This set (η) will automatically be a solution of (1).

We now write down an extended adjoint set of equations

$$\frac{d}{dx}(p_{ki}\lambda_k) - q_{ki}\lambda_k = 0, \quad (i, k = 1, \cdots, n). \tag{8}$$

We will use later the fact that these equations have a continuous

* For a proof of the possibility of this adjunction, see Bliss, loc. cit., p. 312.

solution $(\lambda_1, \cdots, \lambda_n)$ such that the functions $p_{ki}\lambda_k$ take on arbitrary values at one end point (since $|p_{ki}|$ is not zero).

LEMMA. *Between functions $\eta_i(x)$ of class C' satisfying* (7) *with the corresponding ζ's and continuous solutions λ_k of* (8) *we have the following relation*:

$$[p_{ki}^{s}\lambda_k^{s}\eta_i^{s}]_1^2 = \int_{x^1}^{x^2} \lambda_r\zeta_r dx. \tag{9}$$

For it is easily verified that the relation

$$\frac{d}{dx}(p_{ki}\lambda_k\eta_i) - \lambda_r\zeta_r = 0, \quad (i, k = 1, 2, \cdots n; \tau = m + 1, \cdots, n),$$

holds, and we need only integrate this to obtain (9).

THEOREM 2. *A necessary and sufficient condition that the adjoint system* (2), (6) *have a continuous solution $(\lambda_1, \cdots, \lambda_m)$ not zero at any point of (x^1, x^2) is that all determinants of the form*

$$|\eta_{ij}^{s} - \bar{\eta}_{ij}^{s}|, \qquad \begin{matrix}(2n \text{ columns} : j = 1, \cdots, 2n),\\ (2n \text{ rows} : i = 1, \cdots, n; s = 1, 2),\end{matrix} \tag{A}$$

vanish, where $\eta_{i1}^{s}, \cdots, \eta_{i2n}^{s}$ are the end values of $2n$ sets of solutions of class C' of the original differential equations (1) *and $\bar{\eta}_{i1}^{s}, \cdots, \bar{\eta}_{i2n}^{s}$ are $2n$ sets of values which satisfy the boundary conditions* (4) *with $2n$ sets of values of the parameters (u).*

To prove the necessity of the condition, assume that there is a continuous solution $(\lambda_1, \cdots, \lambda_m)$ of (2) and (6) not zero at any point of (x^1, x^2). Then the set $(\lambda_1, \cdots, \lambda_m, 0, \cdots, 0)$ is a solution of (8) and the preceding lemma proves that

$$[p_{\beta i}^{s}\lambda_\beta^{s}\eta_i^{s}]_1^2 = 0. \tag{10}$$

If we multiply equations (6) respectively by arbitrary constants $u_1, \cdots, u_r$ and add, and subtract from (10), we obtain

$$[p_{\beta i}^{s}\lambda_\beta^{s}(\eta_i^{s} - c_{ik}^{s}u_k)]_1^2 = 0, \tag{11}$$

which may be written as

$$[p_{\beta i}^{s}\lambda_\beta^{s}(\eta_i^{s} - \bar{\eta}_i^{s})]_1^2 = 0. \tag{12}$$

The $2n$ coefficients $p_{\beta_i}^1\lambda_\beta^1$, $p_{\beta_i}^2\lambda_\beta^2$ are not all zero, for otherwise we would have $\lambda_\beta^s = 0$, which is contrary to hypothesis. Hence all determinants of form (A) must vanish, and the condition of our theorem is necessary.

Now suppose that all determinants of form (A) vanish. Then there must exist constants m_i^s not all zero such that the equation

$$m_i^s(\eta_i^s - \bar{\eta}_i^s) = 0 \tag{13}$$

holds for all solutions $\eta_i(x)$ of class C' of (1), and all solutions $\bar{\eta}_i^s$ of (4). This is proved as in Morse and Myers (loc. cit., p. 243). Using (4), we may write this in the form

$$\overset{s}{m}_i(\overset{s}{\eta}_i - \overset{s}{c}_{ih}u_h) = 0. \tag{14}$$

Let $(\lambda_1, \cdots, \lambda_n)$ be any continuous solution of (8) not identically zero. Then $(\lambda_1, \cdots, \lambda_n)$ cannot all be zero at any one point of (x^1, x^2) because of the homogeneity of (8). Subtract (9) from (14). We obtain

$$\eta_i^2(m_i^2 - p_{ki}^2\lambda_k^2) + \eta_i^1(m_i^1 + p_{ki}^1\lambda_k^1) - \overset{s}{c}_{ih}u_h\overset{s}{m}_i + \int_{x^1}^{x} \lambda_\tau\zeta_\tau dx = 0, \tag{15}$$

$$(i, k = 1, 2, \cdots, n; h = 1, 2, \cdots, r; \tau = m+1, \cdots, n).$$

According to the sentence following (8), we may consider $p_{ki}^1\lambda_k^1$ as having arbitrary values. Furthermore, according to the remarks immediately preceding (8), we may consider ζ_τ as arbitrary continuous functions of x, and η_i^2 as having arbitrary values. The values of the parameters u_h are also arbitrary.

Let us choose $p_{ki}^1\lambda_k^1 = -m_i^1$. Then if we choose temporarily $\eta_i^2 = 0$ and $u_h = 0$, we find that $\lambda_\tau \equiv 0$. Equation (15) may now be written in the form

$$\overset{2}{\eta}_i(\overset{2}{m}_i - \overset{2}{p}_{ki}\overset{2}{\lambda}_k) - \overset{s}{c}_{ih}u_h\overset{s}{m}_i = 0. \tag{16}$$

From the arbitrariness of u_h and η_i^2 we deduce that

$$\overset{2}{p}_{ki}\overset{2}{\lambda}_k = \overset{2}{m}_i, \quad \overset{s}{m}_i\overset{s}{c}_{ih} = 0.$$

Thus the conditions

$$[\overset{s}{p}_{ki}\overset{s}{\lambda}_k\overset{s}{c}_{ih}]_1^2 = 0, \quad \frac{d}{dx}(p_{ki}\lambda_k) - g_{ki}\lambda_k = 0$$

are satisfied with $\lambda_\tau \equiv 0$, $(\tau = m+1, \cdots, n)$, or, in other words, conditions (6) and equations (2) have a continuous solution not zero. Hence the condition of our theorem is sufficient.

4. *The Problem of Mayer.* Let us consider arcs $y_i = y_i(x)$ of class C' in $(n+1)$-space, let us denote the end points of these arcs by $(x^s, y_1^s, \cdots, y_n^s)$ (x^s, y^s), where $s=1$ at the initial end point and $s=2$ at the final end point.

Such arcs which satisfy the differential equations*

$$\phi_\beta(x, y, y') = 0 \tag{17}$$

are called *differentially admissible* arcs. The functions ϕ_β are of class C''. Arcs of the same type which satisfy, for some value of the parameters $(\alpha) = (\alpha_1, \cdots, \alpha_r)$, the conditions

$$x^s = x^s(\alpha), \quad y_i^s = y_i^s(\alpha), \tag{18}$$

will be called *terminally admissible* arcs. The end point functions $x^s(\alpha)$, $y_i^s(\alpha)$ are to be of class C'. An arc which is both differentially and terminally admissible will be called *admissible.* A set (α) will be called *admissible* if it determines through (18) the end points of some admissible arc.

Let E be an admissible arc $y_i = \bar{y}_i(x)$, $a^1 \leqq x \leqq a^2$, which satisfies (18) for $(\alpha) = (0)$ and along which the functional matrix $\|\phi_{\beta y_i'}\|$ is of rank m. *We seek conditions under which E and the set $(\alpha) = (0)$ afford a minimum for α_1 among admissible sets (α) and the corresponding admissible arcs.*

This simple form of the problem is new. It produces symmetrical results and proofs, and makes clear certain adjoint relationships in the problem. It is as general as any other form previously used in the sense that any of the other forms can be put into this form by simple transformations. The general form used by Morse and Myers (loc. cit.), in which the expression to be minimized is

$$\int_{x^1}^{x^2} f(x,y,y')dx + \theta(\alpha),$$

can be put into the form used in this paper by introducing a new dependent variable y_0 and a parameter α_0, and adjoining the new differential equation $y_0' - f(x, y, y') = 0$, and the new end

* We use the notation (x, y, y') for $(x, y_1, \cdots, y_n, y_1', \cdots, y_n')$.

conditions $y_0^1 = 0$, $y_0^2 = \alpha_0 - \theta(\alpha)$. The expression to be minimized becomes simply α_0.

5. *The Euler-Lagrange Equations and the Transversality Condition.* A set of functions $\eta_i(x)$ of class C' which satisfy the *differential equations of variation*

$$\phi_{\beta y_i'}\eta_i' + \phi_{\beta y_i}\eta_i = 0, \qquad (x, y, y') \text{ on } E, \tag{19}$$

will be called a set of *differentially admissible variations.*

We shall attach the subscript h to $x^s(\alpha)$ or $y_i^s(\alpha)$ to denote differentiation with respect to α_h. Let us set

$$\overset{s}{c}_{ih} = \overset{s}{y}_{ih} - \overset{s}{y}_i' \overset{s}{x}_h, \qquad (h = 1, \cdots, r).$$

Then a set of $2n$ numbers η_i^s which satisfy the *terminal equations of variation*

$$\overset{s}{\eta}_i = \overset{s}{c}_{ih} u_h \tag{20}$$

for some set of numbers $(u) = (u_1, \cdots, u_r)$ will be called a set of *terminally admissible variations* and will be said to be determined by the set (u).

Let $\|\eta_{ij}(x)\|$, $(i = 1, \cdots, n;\ j = 1, \cdots, 2n+2)$, be any matrix consisting of $2n+2$ columns of differentially admissible variations. Let γ_j^1, γ_j^2 be the jth of $2n+2$ sets of 2 arbitrary constants. Then the following theorem can be proved.*

THEOREM 3. *There exists a $(2n+2)$-parameter family of differentially admissible arcs*

$$\begin{gathered} y_i = y_i(x, e_1, \cdots, e_{2n+2}), \\ x_0^1(e_1, \cdots, e_{2n+2}) \leqq x \leqq x_0^2(e_1, \cdots, e_{2n+2}), \end{gathered} \tag{21}$$

containing the arc E for $(e_1, \cdots, e_{2n+2}) = (0, \cdots, 0)$, and such that

$$\overset{s}{x}_{0e_j}(0, \cdots, 0) = \overset{s}{\gamma}_j, \quad y_{ie_j}(x, 0, \cdots, 0) = \eta_{ij}(x), \qquad (j = 1, \cdots, 2n+2).$$

The functions y_i, y_{ix}, and x_0^s are of class C' in $e_1, \cdots, e_{2n+2}$.

* See Morse and Myers, loc. cit., p. 239.

Now let u_{qj}, $(q=2, \cdots, r;\ j=1, \cdots, 2n+2)$, be $2n+2$ sets of $r-1$ arbitrary constants. Consider the equations

$$
\begin{aligned}
&x_0^s(e_1, \cdots, e_{2n+2}) - x^s(\alpha_1, e_j u_{2j}, \cdots, e_j u_{rj}) = 0, \\
&y_i[x_0^s(e_1, \cdots, e_{2n+2}), e_1, \cdots, e_{2n+2}] - y_i^s(\alpha_1, e_j u_{2j}, \cdots, e_j u_{rj}) = 0.
\end{aligned} \tag{22}
$$

These equations are conditions for terminal admissibility and they have the initial solution $(\alpha_1, e_1, \cdots, e_{2n+2}) = (0, 0, \cdots, 0)$. If the jacobian of the left hand sides of equations (22) with respect to $e_1, \cdots, e_{2n+2}$ were different from zero at this solution, the implicit function theorems would show that (22) would have solutions for $(e_1, \cdots, e_{2n+2})$ as functions of α_1 for all values of α_1 in the neighborhood of $\alpha_1 = 0$. That is, there would be admissible arcs in the family (21) giving α_1 a value smaller than zero. This would contradict the fact that E is a minimizing arc. Hence we conclude that the jacobian must be zero.

If we attach the subscript p to x^s or y_i^s to represent differentiation with respect to α_p, the jacobian in question is

$$
\left| \begin{array}{c} \gamma_j^s - x_p^s u_{pj} \\ \\ \bar{y}_i^{rs}\gamma_j^s + \eta_{ij}^s - y_{ip}^s u_{pj} \end{array} \right| \begin{array}{r} (p = 2, \cdots, r), \\ (2n+2 \text{ columns};\ j = 1, \cdots, 2n+2), \\ (2n+2 \text{ rows};\ s = 1, 2;\ i = 1, \cdots, n). \end{array} \tag{23}
$$

Let us set $\breve{\gamma}_j^s = x_p^s u_{pj}$, $\breve{\eta}_{ij}^s = c_{ip}^s u_{pj}$, $(p=2, \cdots, r)$. If we multiply the first two rows of the Jacobian respectively by $\bar{y}_i^{r1}$ and $\bar{y}_i^{r2}$ and subtract them respectively from the next n rows and the last n rows, we obtain

$$
\left| \begin{array}{c} \gamma_j^s - \breve{\gamma}_j^s \\ \eta_{ij}^s - \breve{\eta}_{ij}^s \end{array} \right| = 0. \tag{24}
$$

This holds for arbitrary constants γ_j^s, for any sets of differentially admissible variations $\eta_{ij}(x)$, and for arbitrary constants u_{pj}.

The arbitrariness of γ_j^s enables us to deduce from (24) the following necessary condition:

$$
\left| \eta_{i\delta}^s - \breve{\eta}_{i\delta}^s \right| = 0, \qquad (\delta = 1, \cdots, 2n), \tag{25}
$$

where $\eta_{i\delta}(x)$ is the δth of any $2n$ sets of differentially admissible

variations, and where $u_{p\delta}$ is arbitrary, $\overset{s}{\eta}_{i\delta}$ being given by the following:

$$\overset{s}{\eta}_{i\delta} = \overset{s}{c}_{ip}u_{p\delta}, \qquad (p = 2, \cdots, r). \tag{26}$$

But, according to Theorem 2 of §3, equation (25) is a sufficient condition that the following adjoint system of (20) and (26) have a continuous solution $(\lambda_1, \cdots, \lambda_m)$ not zero at any point of (a^1, a^2):

$$\frac{d}{dx}(\phi_{\beta y_i'}\lambda_\beta) - \phi_{\beta y_i}\lambda_\beta = 0, \quad (\beta = 1, \cdots, m;\ i = 1, \cdots, n), \tag{27}$$

$$[\overset{s}{\phi}_{\beta y_i'}\overset{s}{\lambda}_\beta \overset{s}{c}_{ip}]_1^2 = 0, \qquad (p = 2, \cdots, r). \tag{28}$$

But from (28) follows an identity in the differentials $d\alpha_p$;

$$[\overset{s}{\phi}_{\beta y_i'}\overset{s}{\lambda}_\beta \overset{s}{c}_{ip}]_1^2 d\alpha_p = 0, \qquad (p = 2, \cdots, r).$$

Now $c_{ih}^i d\alpha_h = dy_i^s - \bar{y}_i'^s dx^s$, $(h = 1, 2, \cdots, r)$, and so we obtain

$$[\overset{s}{\phi}_{\beta y_i'}\overset{s}{\lambda}_\beta(d\overset{s}{y}_i - \overset{s}{\bar{y}'}_i\, d\overset{s}{x})]_1^2 + K d\alpha_1 = 0, \ (K = \text{constant}). \tag{29}$$

These results may be summarized in the following theorem.

THEOREM 4. *If E and the set $(\alpha) = (0)$ afford a minimum in the problem, there exist a constant K and a set of continuous functions $\lambda_1(x), \cdots, \lambda_m(x)$, not all zero at any point of (a^1, a^2), such that if we set $F = \lambda_\beta\phi_\beta$, the equations*

$$\frac{d}{dx}F_{y_i'} - F_{y_i} = 0 \tag{30}$$

are satisfied at every point of E, and the following identity in $(d\alpha)$ holds:

$$[\overset{s}{F}_{y_i'}(d\overset{s}{y}_i - \overset{s}{\bar{y}}_i d\overset{s}{x}]_1^2 + K d\alpha_1 = 0, \tag{31}$$

when dx^s and dy_i^s are expressed in terms of $(d\alpha)$ and evaluated for $(\alpha) = (0)$.

6. *Normalcy.* An admissible arc E is said to be *normal relative to conditions* (30) *and* (31) if there exists no set of continuous functions $\lambda_\beta \not\equiv 0$ with which it satisfies these conditions with $K = 0$; that is, if it does not satisfy the differential equations and boundary conditions

$$\frac{d}{dx}(\phi_{\beta y_i'}\lambda_\beta) - \phi_{\beta y_i}\lambda_\beta = 0, \tag{32}$$

$$[\overset{s}{\phi}_{\beta y_i'}\overset{s}{\lambda}_\beta \overset{s}{c}_{ik}]_1^2 = 0, \qquad (h = 1, \cdots, r). \tag{33}$$

Otherwise E is said to be *anormal* relative to these conditions.

From Theorems 1 and 2 the following theorems are evident.

THEOREM 5. *The system* (32) *and* (33) *is an adjoint system of the differential and terminal equations of variations* (19) *and* (20).

THEOREM 6. *A necessary and sufficient condition for normalcy relative to* (30) *and* (31) *is that there exist a non-vanishing determinant of the form*

$$|\overset{\delta}{\eta}_{is} - \overset{\delta}{\bar{\eta}}_{is}|, \quad (\delta = 1, \cdots, 2n;\ i = 1, \cdots, n;\ s = 1, 2),$$

where $\eta_{is}(x)$, $(i=1, \cdots, n)$, *is the* δ*th of any* $2n$ *sets of differentially admissible variations, and* $\bar{\eta}^s_{is}$ *is the* δ*th of any* $2n$ *sets of terminally admissible variations.*

HARVARD UNIVERSITY

OFFICIAL COMMUNICATIONS

Meetings of the Society have been fixed at the following times and places:

On account of the fact that the Summer Meeting of the Society in 1932 will be held on the Pacific Coast, the meetings that were to have been held at Stanford University (April, 1932), at Seattle (June, 1932), and at Los Angeles (November, 1932) will be omitted.

LOS ANGELES, CALIFORNIA, SUMMER MEETING AND COLLOQUIUM, August 30-September 2, 1932.

Abstracts must be in the hands of Associate Secretary M. H. Ingraham, University of Wisconsin, Madison, Wis., not later than July 26. Abstracts received by July 7 will appear in the July issue, or an earlier issue, of this BULLETIN. Professor J. F. Ritt will deliver a series of colloquium lectures on *Differential equations from the algebraic standpoint.* By invitation of the program committee, Professor D. N. Lehmer will speak on *The continued fraction representing cubic and higher irrationalities,* and Professor Tibor Rado will speak on *Recent work in the problem of Plateau.*

NEW YORK CITY, October 29, 1932.

Abstracts must be in the hands of Associate Secretary Tomlinson Fort, 501 West 116th St., New York City, not later than October 8. In order to be printed in the BULLETIN in advance of the meeting, abstracts must be in the hands of the Associate Secretary not later than September 7. By invitation of the program committee, Professor Jesse Douglas will speak on *The problem of Plateau.*

ATLANTIC CITY, NEW JERSEY, ANNUAL MEETING, December 27-30, 1932.

Abstracts are due by November 30, 1932.

R. G. D. RICHARDSON, *Secretary of the Society.*

Articles for insertion in the BULLETIN should be addressed to E. R. HEDRICK, Editor of the BULLETIN, University of California at Los Angeles. Reviews should be sent to W. R. LONGLEY, Yale University, New Haven, Conn. Notes should be sent to H. W. KUHN, Ohio State University, Columbus, Ohio.

Subscriptions to the BULLETIN, orders for back numbers, and inquiries in regard to non-delivery of current numbers should be addressed to the American Mathematical Society, 450-459 Ahnaip St., Menasha, Wis., or 501 West 116th St., New York.

The initiation fees and the annual dues of members of the Society (see this BULLETIN, p. 322, May, 1930; and the List of Officers and Members, October, 1930, p. 58), are payable to the Treasurer of the Society, Professor G. W. Mullins, 501 West 116th St., New York City.

WHOLE NUMBER 391

CONTENTS

PAGE

The Calculus of Variations and the Quantum Theory. By G. A. BLISS 201
Statistical Mechanics and the Second Law of Thermodynamics. By P. W. BRIDGMAN 225
Note on Diophantine Automorphisms. By LEONARD CARLITZ 246
On a Matrix Differential Operator. By A. K. MITCHELL 251
Solutions of Bounded Variation of the Fredholm-Stieltjes Integral Equation. By L. J. PARADISO 255
On Real Symmetric Determinants whose Principal Diagonal Elements are Zero. By W. V. PARKER 259
On the Degree of Generality of a Class of Arithmetical Identities. By E. T. BELL 263
On the Moment Problem for a Finite Interval. By T. H. HILDEBRANDT 269
New Diophantine Automorphisms. By LEONARD CARLITZ 271
Note on the Diophantine Equation. $ax^2+by^2+cz^2+dt^2=0$. By J. L. MORDELL 277
Note on an Application of Metric Geometry to Determinants. By L. M. BLUMENTHAL 283
On Polynomial Interpolation to Analytic Functions with Singularities. By J. L. WALSH 289
Analytic Study of Rational Quintic Surfaces having no Multiple Curves. By H. N. HUBBS 295
Adjoint Systems in the Problem of Mayer under General End-Conditions. By S. B. MYERS 303

For official Communications and Notices, see the inside of the back cover.

GEORGE BANTA PUBLISHING COMPANY, MENASHA, WISCONSIN